1-1-2015

Development Of New Algorithms For Exploring The Potential Energy Landscape Of Chemical Reactions

Adam Benjamin Birkholz
Wayne State University,

DEVELOPMENT OF NEW ALGORITHMS FOR EXPLORING THE POTENTIAL ENERGY LANDSCAPE OF CHEMICAL REACTIONS

By

ADAM BENJAMIN BIRKHOLZ

DISSERTATION

Submitted to the Graduate School

of Wayne State University

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2015

MAJOR: CHEMISTRY (Physical)

Approved By:

Advisor Date

ACKNOWLEDGMENTS

I would like to thank, first and foremost, Professor H. B. Schlegel. Without his patience and support over the years, this thesis would not have been possible. I am also grateful to Professor Schlegel and the other members of my committee, Professors Andres Cisneros, Vladimir Chernyak and Robert Reynolds for the guidance, discussion and suggestions that have helped to shape the work presented in this thesis, while also encouraging the growth and development of my present and future scientific career. I also owe a great deal to Rebecca Swett, as this thesis would not have been possible without her constant and unwavering friendship, support, and advice over the years since we started grad school together.

I have shared the Schlegel group office with a number of different graduate students and post docs over the past 7 years, and I am grateful for the many opportunities I've had to learn from and collaborate with them. Thank you Doctors Barb Munk, Jason Sonnenberg, Richard Lord, Sue Li, Jia Zhou, Jason Sonk, Brian Psciuk, Pascal Krause, Shivnath Mazumder, Qing Liao, and current phd students Xuetao Shi, Bishnu Thapa, Sebastian Herbert and Yi-Jung Tu. I would also like to thank the past and present members of the Cisneros and Chernyak groups, for giving me an opportunity to become familiar with the work being done in their labs, and for providing invaluable outside insight and discussion about my work during our joint group meetings.

Wayne State Chemistry's amazing support staff played a huge role in helping me navigate the technical and bureaucratic hurdles involved with being a PhD student. Melissa Barton, Mary Wood, Bernie Miesik, Nestor Ocampo and all of the other front office and business office staff that have helped me in one way or another all deserve my sincerest thanks. The financial contributions from the NSF, Wayne State University and Gaussian Inc are greatly appreciated for the role they played in supporting my

work.

TABLE OF CONTENTS

Acknowledgements . . . . . **ii**

List of Figures . . . . . **vii**

List of Tables . . . . . **x**

1 Introduction **1**

1.1 Coordinate Choice for Exploring Chemical Reaction Paths 3

1.2 Improvements to Optimization Methods 4

1.3 Development of a Variational Reaction Coordinate Method 6

2 Coordinate Reduction for Exploring Chemical Reaction Paths . . . 10

2.1 Introduction 10

2.2 Methods 12

2.3 Discussion and Results 15

2.4 Conclusions 19

3 Using Bonding to Guide Transition State Optimization 29

3.1 Introduction 29

3.2 Methods 32

3.2.1 Choice of Coordinate System 32

3.2.2 Bond Order Interpolation with Relaxation 34

3.2.3 Using Connectivity Change to Approximate the Transition Vector 38

3.2.4 Initial approximation of the Hessian 41

3.2.5 Divided RFO 44

3.3 Implementation and Discussion of Results 45
3.4 Summary 48
4 Exploration of Improvements to Optimization Methods 55
4.1 Introduction 55
4.2 Standard Method 56
4.2.1 Coordinate transformation/projection 57
4.2.2 Linear Bends 62
4.2.3 Step computation/constraint 65
4.2.4 Hessian Update 67
4.2.5 Convergence 68
4.3 Flowchart Update 68
4.3.1 Motivation 68
4.3.2 Method 68
4.4 Scaled RFO method 70
4.4.1 Motivation 70
4.4.2 Method 71
4.5 Quasi-rotation internal coordinate propagation 72
4.5.1 Motivation 72
4.5.2 Method 73
4.6 Results and Discussion 74
5 Path Optimization by a Variational Reaction Coordinate Method . 80
5.1 Introduction 80

5.2 Path Optimization by VRC 82

5.2.1 A continuous description of the reaction path 82

5.2.2 VRE derivatives 85

5.2.3 Constraints and constrained optimization 89

5.3 Combined path and transition state optimization 96

5.3.1 Coupling constraints 96

5.3.2 Geometry optimization 100

5.3.3 Handling rotations 101

5.3.4 Multiple Extrema 103

5.3.5 FVRC algorithm 104

5.4 Results and Discussion 105

5.5 Conclusions and future directions 107

6 Incorporating Redundant Internal Coordinates into the Variational Reaction Coordinate Method 116

6.1 Introduction 116

6.2 RIC Coupling Constraints for the FVRC Method 120

6.3 Least Length Path 122

6.3.1 RIC Arc Length 122

6.3.2 Derivatives and Modified RIC Length Lagrangian 125

6.3.3 Improved $\boldsymbol{\lambda}_{\boldsymbol{\sigma}}$ Initialization 127

6.3.4 RIC Arc Length Minimization Algorithm 129

6.4 Redundant Internal Coordinate VRC method 130

6.4.1 RIC VRE definition 130
6.4.2 Approximating the RIC PES 131
6.4.3 RIC-FVRC algorithm 134
6.5 Results and Discussion 135
7 Summary and Future Directions 142
References 145
Abstract 154

LIST OF FIGURES

Figure 1.1 An example potential energy surface with an indication of some of the features studied by computational chemistry 1

Figure 2.1 Representative test reactions 20

Figure 2.2 Additional test reactions 21

Figure 2.3 Reaction path for the ene reaction. **a** The three Cartesian coordinates with the largest variance and the Cartesian coordinate with the 10th largest variance. **b** The three Cartesian principal components with the largest variance and the Cartesian principal component with the 10th largest variance. 22

Figure 2.4 Energy along the reaction path for the ene reaction using the first d principal components to represent the reaction path, $d = 1 - 4$. **a** Cartesian coordinates. **b** Redundant internal coordinates 23

Figure 2.5 Mean absolute error and maximum absolute error in the energy along the path for the ene reaction as a function of the number of principal components, the dashed horizontal line indicates chemical accuracy of 1 kcal/mol 24

Figure 2.6 Mean absolute error and maximum absolute error in the Cartesian coordinates along the path for the ene reaction as a function of the number of principal components. The dashed horizontal line indicates chemical accuracy of 0.005 angstroms 25

Figure 2.7 Contour plot of the mean absolute error in the coordinates along the path for the ene reaction as a function of the number of principal components and the number of images used to generate the principal components, with the darkest region corresponding to an error of 0.005 angstroms. The inset shows the mean absolute error in the computation of the average coordinates versus the number of images in the subset path 26

Figure 3.1 Representative test reactions 50

Figure 4.1 Structures for test set 77

Figure 4.1 Structures for test set (cont) 78

Figure 5.1 B-Spline basis with 5 internal functions ($\phi_2 - \phi_7$) and 2 capping functions (ϕ_1 and ϕ_7) 109

Figure 5.2 Three cubic spline curves fit to the same 5 points, using different values of t. Exact t = (0,0.25,0.5,0.7,1), Shift t = (0,0.15,0.5,0.85,1), Skew t = (0,0.1,0.4,0.7,1) 110

Figure 5.3 Comparison of the shape of the converged unconstrained VRC (UVRC), constrained VRC (CVRC) and focused VRC (FVRC) pathways on the Müller-Brown analytical surface. The solid curve is the IRC, and the large dots are the minima and transition states along the IRC. . . . 111

Figure 5.4 Convergence log plots for the various VRC methods on the Müller-Brown surface. **Top** RMS VRE gradient **Center** RMS PES gradient (measured at the minima and maxima along the path only) **Bottom** RMS LEC displacement 112

Figure 5.5 The paths at every iteration of the FVRC method on the Müller-Brown surface, compared with the IRC. 113

Figure 5.6 Convergence of the VRE and PES gradient using the FVRC method on the LJ10 surface. 114

Figure 5.7 Energy plots for selected iterations of the LJ10 FVRC optimization. **Top** Iterations 1-4 **Center** Iterations 4-7, zoomed in to show detail **Bottom** Comparison of the 7th and final (13th) steps, demonstrating that by the 7th iteration, the path is already in the same region of the PES as the converged path 115

Figure 6.1 Reactant (left) and Product (right) structures. From top to bottom: MeOH, Ene, Cope, Bispidine. For the MeOH and Cope structures, key atoms have been labeled 139

Figure 6.2 Comparison of the maximum energy structures along the path for the Ene (left) and Cope (right) reactions. From top to bottom, Cartesian interpolation (XLST), least length RIC path, RIC-CTS path, final converged path. 140

Figure 6.3 Comparison of energy profiles for the cartesian (XLST), least length RIC (RIC), RIC-CTS and final converged path. **a** MeOH, **b** Ene, **c** Cope, **d** Bispidine. The plot for Bispidine is zoomed in to show the fine structure of the RIC and final paths, the maximum along the XLST path is approximately 0.5 hartree. 141

LIST OF TABLES

Table 2.1 Comparison of the number of principal components needed to achieve chemical accuracy (<1 kcal/mol) in the mean and maximum absolute error in the energy of the reduced path 27

Table 2.2 Comparison of the number of principal components needed to achieve chemical accuracy (0.005 angstroms) in the mean and maximum absolute error in the Cartesian coordinates of the reduced path 28

Table 3.1 Barrier distances versus the bond order at the PM6 transition state geometry 51

Table 3.2 Comparison of overlap between different tangent approximations and eigenvectors of the Hessian computed at the transition state of the PM6 PES. 52

Table 3.3 Number of Surface Evaluations Required to Converge at PM6 level of theory 53

Table 3.4 Number of Surface Evaluations Required to Converge at B3LYP/6-31G(d,p) level of theory 54

Table 4.1 Number of iterations required to converge to a minimum structure 79

CHAPTER 1

INTRODUCTION

The exploration of potential energy surfaces (PES), which relate the energy of a chemical system to the internal geometry of it's component atoms, is often an important first step in the modern study of reaction mechanisms. The PES can be understood as a landscape with varying elevations representing the energy of the system, with the horizontal directions representing a change in the geometry of the atoms (see above figure). The lowest points located in the valleys of the PES correspond to the reactants, products and intermediates of a reaction, while the maximum point along low-lying hills connecting two nearby valleys can be viewed as the barrier or transition state for the reaction. The pathways that begin in one valley, climb over the transition state, and end in another valley are of particular interest, as these pathways provide a basic description of how the chemical system changes as it undergoes a reaction.

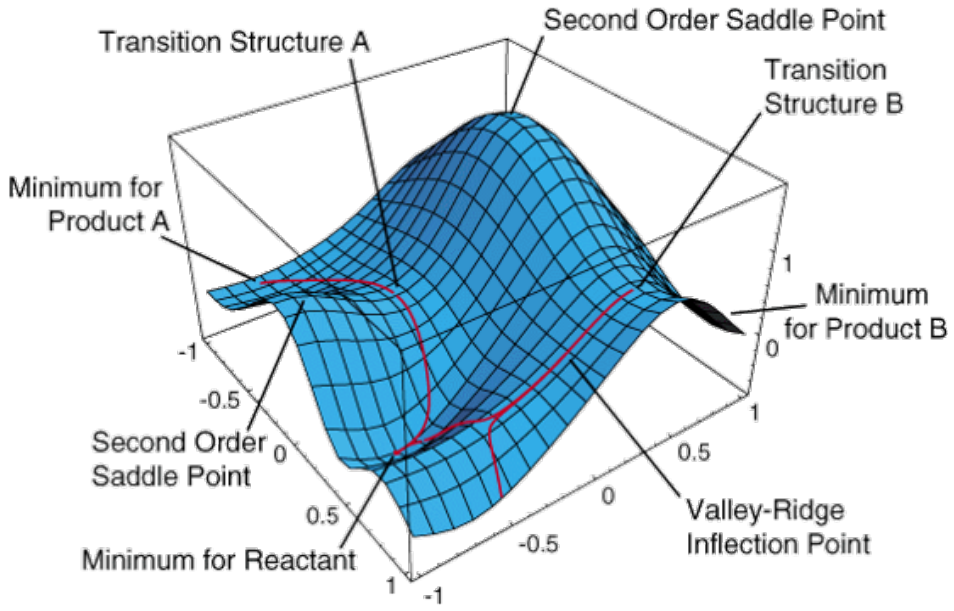


Figure 1.1: An example potential energy surface with an indication of some of the features studied by computational chemistry

Many methods exist to locate minima, transition states, and reaction paths by computing the height (energy), slope (gradient, first derivative) and curvature (Hessian, second derivative) of the PES, and using those values to step closer to the point of interest or to improve the current approximation to a reaction pathway. When high chemical accuracy is required, computationally demanding ab initio electronic structure and density functional theory methods must be used to evaluate the PES. Since these methods take 2-4 orders of magnitude more computational effort to compute as a typical optimization step, the number of times that the energy, gradient and Hessian needs to be evaluated is the dominant factor in determining how long it will take for an optimization to finish. Consequently, this number may be used to compare the performance of different optimization methods without discussing CPU time, which can depend heavily on factors external to the choice of optimization method (such as computer architecture, choice of electronic structure program, and so forth). New methods that can reduce this number by improving the approximation of initial geometries prior to optimization, or by making better use of the information available each time the PES is evaluated, will help to extend the computational study of chemical reactions to new domains that require a combination of higher accuracy or the study of larger molecules than current methods can support on modern computer architectures.

The work presented in this thesis is divided into 5 chapters. Chapters 2 and 3 focus on the effect that the choice of coordinate representation can have on the efficiency and reliability of representing and exploring the PES in the region of transition states. Chapter 4 introduces a variety of new algorithms which may help reduce the computational cost required to locate minimum energy structures. Chapters 5 and 6 outline a new approach to optimizing the reaction path without first locating a transition state. Each of these chapters provide detailed descriptions of the implementation of the relevant algorithms, as well as data to support their usage and motivate further

development. Chapter 7 briefly summarizes the methods for exploring the PES developed in this thesis, and discusses of some of the many possibilities that are available for future research and development.

1.1 Coordinate Choice for Exploring Chemical Reaction Paths

A potential energy surface describes the energy of a molecule as a function of its geometrical parameters[1]. The potential energy surface representing the energetics of an N atom molecule is defined in terms of $3N$ Cartesian coordinates or at least $3N-6$ internal coordinates. Minima on the potential energy surface correspond to equilibrium structures such as reactants and products, and first order saddle points represent transition states for reactions. There are many areas of active study involving the exploration of these surfaces, including geometry optimization[2], reaction path following[3], reaction path optimization[4, 5, 6], potential energy surface interpolation[7, 8], and molecular dynamics[9]. The computational difficulty of these algorithms depends heavily upon how many degrees of freedom are used to define the potential energy surface. Any approach to generate a reduced set of coordinates that adequately describes the region of the surface to be explored could improve the efficiency of these algorithms.

A set of (possibly redundant, i.e. more than 3N-6) internal coordinates comprised of stretches, bends and torsion involving bonded atoms often provides a better chemical description of the structure and flexibility of a molecule than Cartesian coordinates[10, 11]. Delocalized internal coordinates[12] are more compact than primitive redundant internal coordinates and have been employed for geometry optimization and vibrational analysis. They are more general than non-redundant (Z-matrix) internal coordinates while still using $3N-6$ coordinates to represent the potential energy surface, however they may be more strongly coupled than primitive redundant internal coordinates. Reaction paths have been analyzed in terms of Cartesian coordinates,

internal coordinates[13] and adiabatic local modes[14]. Because reactions often involve the making and/or breaking of a small number of bonds, it should be possible to represent the paths with far fewer than $3N - 6$ coordinates. In this chapter, principal component analysis (PCA)[15] is used to obtain a reduced number of coordinates for a reaction path. PCA has been applied successfully to obtain an improved set of internal coordinates for vibrational analysis[16], and to examine conformational changes in molecular dynamics simulations of biomolecules[17, 18].

For a given reaction, one can define a steepest descent reaction path that connects the reactant minimum through the transition state to the products[19]. Reaction path following can be carried out in Cartesian or internal coordinates, with or without mass-weighting[20, 21]. Since a reaction usually involves significant changes in only a few bonds or angles, less than $3N-6$ coordinates should be needed to represent the changes in a molecule along a reaction path. The rest of the coordinates remain approximately constant as the molecule moves along the path from reactants to products. In Chapter 2, PCA is used to determine a well chosen subset of (redundant) internal coordinates that provides a significant reduction in the number of coordinates needed to represent the path. In Chapter 3, the knowledge about how the connectivity changes between reactants and products is used to improve the generation of approximate structures for transition state optimization, as well as to improve the efficiency of the transition state optimization itself.

1.2 Improvements to Optimization Methods

Geometry optimization is an important tool in the computational chemistry toolbox, and has become ubiquitous in modern studies of chemical properties and reactions. There are a wide variety of different algorithms that exist for optimization (see ref [2] for a recent review of methods), with the most common utilizing a combination of quasi-

Newton steps in redundant internal coordinates[10, 11], with the Hessian updated by the method of Broyden Fletcher Goldfarb and Shanno (BFGS[22, 23, 24, 25]) when seeking a minimum structure, and either the Powell's symmetric Broyden (PSB[26]) or the symmetric rank 1 (SR1[27]) updates, or some combination of the two[28], is used when attempting to locate a transition state. Approximate, positive definite Hessian matrices[29] are typically used to avoid the relatively large cost of computing the full Hessian exactly which is often an order of magnitude more costly than computing the energy and the gradient. Additionally, sequence acceleration methods such as line searches and direct inversion of the iterative subspace (GDIIS[30]) are also used to reduce the number of potential energy surface (PES) calculations necessary to converge to the desired minimum or transition state structure.

Chapter 4 describes three new, alternative approaches that may offer additional benefits or improved performance over existing methodologies.

- Flowchart Update - This approach seeks to improve Hessian updating by using different update methods only when they are expected to be well behaved, and falling back to more reliable but less ideal updates when necessary. Additionally, a new modification to the PSB method is used by using scaled displacements to compute the update.

- Scaled RFO method - This approach seeks to improve the use of the rational function optimization method for controlling step size and direction by modifying the shift-matrix to better represent the expected relative stiffness of the bond stretches versus the other coordinates.

- Quasi-rotation method - This is an alternative approach to handling the redundancy in an internal coordinate system. Rather than store the approximate Hessian in the full redundant space, a quasi-rotation matrix is used to rotate

the approximate Hessian from the non-redundant space at one point, to the non-redundant space at another. Ideally, this will lead to a more consistent approximation to the Hessian even when the non-redundant space changes over the course of the optimization, and may help improve Hessian updating since the change in the gradient can be expressed entirely in the non-redundant space at one set of coordinates.

1.3 Development of a Variational Reaction Coordinate Method

With recent developments in reaction path following[19, 20, 31], the steepest descent reaction path (SDRP) can be readily determined by walking downhill from a transition state on the potential energy. At any point along this pathway, the energy orthogonal to the path is a local minimum, and so it describes a first-order description of the route that a chemical system follows as a reaction proceeds from reactants to products. Locating the transition state in order to determine the SDRP, however, can often be a difficult task. One common approach to approximate the minimum energy path without a converged transition state structure is to express the pathway as multiple points which are optimized simultaneously. These "chain of states" methods typically begin with a series of images along the linear interpolation between reactants and products[32], after overall translation and rotation have been removed. These images are updated to minimize the energies of each point subject to constraints, fictitious forces or interpolation/reparameteration schemes, which ensure that the points maintain a uniform description of the pathway. In these methods, the optimizer is generally required to take small steps in order to avoid the introduction of kinks in the path due to the discrete representation of the path. This ad hoc approach also has the draw back that it is not variational, so there is no reliable way of determining whether or not the optimization is making good progress, or if a solution found is in

fact a minimum.

The line integral of the gradient norm is variational property of a reaction path[33, 34], and can be computed as

$$\begin{aligned} E_{VRE} &= \int_{t_R}^{t_P} \sqrt{\frac{\partial V\left(\mathbf{x}\left(t\right)\right)}{\partial x}^{\mathrm{T}} \frac{\partial V\left(\mathbf{x}\left(t\right)\right)}{\partial x}} \sqrt{\frac{d\mathbf{x}\left(t\right)}{dt}^{\mathrm{T}} \frac{d\mathbf{x}\left(t\right)}{dt}} dt \\ &= \int_{t_R}^{t_P} \left|\mathbf{g}\left(\mathbf{x}\left(t\right)\right)\right| \left|\boldsymbol{\tau}\left(t\right)\right| dt \end{aligned} \tag{1.1}$$

where V is the potential energy, $\mathbf{x}\left(t\right)$ are the coordinates of the reaction path parameterized by t, t_R and t_P are the parameter values corresponding to the reactant and product structures, respectively, while $\mathbf{g}$ and $\boldsymbol{\tau}$ are used as shorthand for the gradient of the potential, and the tangent to the path. This integral is a non-negative, energetic quantity, and will be referred to as the Variational Reaction Energy (VRE) throughout this work. Rigorous proofs that the VRE is minimized by the steepest descent reaction path are discussed in work by Quapp, Bofill and others[33, 34], but a simple proof is provided by computing the VRE assuming that $\mathbf{x}\left(t\right)$ is the steepest descent path. This implies that the tangent to the path is everywhere proportional to the gradient, which simplifies the VRE to the absolute value of the projection of the gradient onto the tangent *(E_{pVRE})*

$$\begin{aligned} E_{pVRE} &= \int_{t_R}^{t_P} \left|\mathbf{g}\left(\mathbf{x}\left(t\right)\right)^{\mathrm{T}} \boldsymbol{\tau}\left(t\right)\right| dt \\ &= \sum_a \left(2V\left(\mathbf{x}_{a,TS}\right) - V\left(\mathbf{x}_{a,P}\right) - V\left(\mathbf{x}_{a,R}\right)\right) \end{aligned} \tag{1.2}$$

where the sum on a is over the number of barriers along the path, and $\mathbf{x}_{a,R}$ and $\mathbf{x}_{a,P}$ are the local minimum structures adjacent to the corresponding local maximum

$\mathbf{x}_{a,TS}$. Since the tangent cannot be changed anywhere along the path without removing the proportionality with the gradient, and since the SDRP steps down from the reaction pathway in the direction that makes the largest change to the magnitude of the gradient, the SDRP must minimize the VRE. While the projected VRE will only equal the VRE when $\mathbf{x}(t)$ is the SDRP, it may be computed for any arbitrary path connecting the reactants and products. This provides a useful non-negative estimate for the variational error in the current path

$$\begin{aligned} \epsilon &= E_{VRE} - E_{pVRE} \\ &= \int_{t_R}^{t_P} |\mathbf{g}(\mathbf{x}(t))| \, |\boldsymbol{\tau}(t)| \, dt - (2V(\mathbf{x}_{TS}) - V(\mathbf{x}_P) - V(\mathbf{x}_R)) \end{aligned} \tag{1.3}$$

Methods to minimize E_{VRE} by a chain of states approach have been discussed elsewhere[33, 34], but these suffer from many of the same problems that exist in the ad hoc path optimization methods. Many small steps are required to converge and discretization error can result in non-variational behavior unless many images are used. Since E_{VRE} is a functional of a smooth, continuous object, it should be advantageous to describe the path using a continuous representation such as a basis set expansion. Such a representation provides a set of coordinates, the linear expansion coefficients (LEC), which can be optimized by minimization of E_{VRE} using standard gradient-based optimization methods.

Chapter 5 explores the challenges and concerns with developing such an algorithm. The resulting Focused VRC method is able to determine both a good approximation to the SDRP between two minimum structures, as well as the fully converged geometries of any transition states and intermediates along the SDRP. This strong algorithmic efficiency comes at the unfortunately high per-iteration cost resulting from the need to determine the gradient and Hessian of E_{VRE} with respect to a change in the LEC

by numerical quadrature methods. Chapter 6 explores three ways to incorporate redundant internal coordinate information into the FVRC method in order to reduce the per-iteration cost and make the method more competitive with the ad hoc chain of states approaches:

- Express the FVRC coupling constraints in terms of redundant internal coordinate differences in order to avoid having to deal with a separate rotational alignment step
- Applying the methods developed for the CVRC method to optimize a least RIC length pathway by minimizing the arc length expressed in RIC. The resulting path should be a better initial guess for the VRC method than a linear Cartesian pathway
- Define the VRE and it's derivatives in terms of an interpolated RIC PES.

CHAPTER 2

COORDINATE REDUCTION FOR EXPLORING CHEMICAL REACTION PATHS

Reproduced with permission from *Theor. Chem. Acc.* **2012**, 131, 1-8

2.1 Introduction

A potential energy surface describes the energy of a molecule as a function of its geometrical parameters[1]. The potential energy surface representing the energetics of an N atom molecule is defined in terms of $3N$ Cartesian coordinates or at least $3N-6$ internal coordinates. Minima on the potential energy surface correspond to equilibrium structures such as reactants and products, and first order saddle points represent transition states for reactions. There are many areas of active study involving the exploration of these surfaces, including geometry optimization[2], reaction path following[3], reaction path optimization[4, 5, 6], potential energy surface interpolation[7, 8], and molecular dynamics[9]. The computational difficulty of these algorithms depends heavily upon how many degrees of freedom are used to define the potential energy surface. Any approach to generate a reduced set of coordinates that adequately describes the region of the surface to be explored could improve the efficiency of these algorithms.

A set of (possibly redundant) internal coordinates comprised of stretches, bends and torsion involving bonded atoms may provide a better chemical description of the structure and flexibility of a molecule than Cartesian coordinates[10, 11]. Delocalized internal coordinates[12] are more compact than primitive redundant internal coordinates and have been employed for geometry optimization and vibrational analysis. They are more general than non-redundant (Z-matrix) internal coordinates, but still use $3N-6$ coordinates to represent the potential energy surface. Reaction paths have been analyzed in terms of Cartesian coordinates, internal coordinates[13] and adia-

batic local modes[14]. Because reactions often involve the making and/or breaking of a small number of bonds, it should be possible to represent the paths with far fewer than $3N-6$ coordinates. In this chapter, principal component analysis (PCA)[15] is used to obtain a reduced number of coordinates for a reaction path. PCA has been applied successfully to obtain an improved set of internal coordinates for vibrational analysis[16], and to examine conformational changes in molecular dynamics simulations of biomolecules[17, 18].

For a given reaction, one can define a steepest descent reaction path that connects the reactant minimum through the transition state to the products[19]. Reaction path following can be carried out in Cartesian or internal coordinates, with or without mass-weighting[20, 21]. Since a reaction usually involves significant changes in only a few bonds or angles, less than $3N-6$ coordinates should be needed to represent the changes in a molecule along a reaction path. The rest of the coordinates remain approximately constant as the molecule moves along the path from reactants to products. A well chosen subset of (redundant) internal coordinates may provide a suitable reduction in the number of coordinates needed to represent the path. However, it is often difficult to choose such coordinates manually. PCA can be applied to the geometries that define a reaction path, described by either internal or Cartesian coordinates. This is a systematic method of obtaining a small set of coordinates that can be used to reproduce the reaction path with chemical accuracy.

In the following sections a PCA-based coordinate reduction procedure is described and applied to a number of reactions for which the reaction path has already been computed in order to provide a benchmark for the method's effectiveness in generating a compact description of a reaction path. The ene reaction is used to illustrate the properties and convergence behavior of the coordinate reduction method. The approach is then tested on a set of model systems that are representative of a variety of different

reaction schemes. Finally, the method is applied to a few reactions from the recent literature to demonstrate the coordinate reduction scheme on some larger systems.

2.2 Methods

Consider a reaction path represented by a set of p structures, with index $1 \leq k \leq p$. Each structure can be described by $3N$ Cartesian coordinates, where N is the number of atoms. Let c_μ^k be the Cartesian coordinates, where the superscript indicates the sequence number of the structure along the path. Alternatively, internal coordinates, q_μ^k, can be used to describe the structures. Greek subscripts, $1 \leq \mu \leq m$, are used for the coordinate number. Overall translation and rotation are eliminated from the Cartesian coordinates by moving the structures to a common origin and aligning them to a common axis. Any discontinuities in the torsional coordinates are removed by adding or subtracting 2π as necessary. Linear bends can be avoided by including suitable dummy atoms in the internal coordinate definitions.

$$\bar{x}_\mu = \frac{1}{p}\sum_{l=1}^{p} c_\mu^l \quad \text{or} \quad \bar{x}_\mu = \frac{1}{p}\sum_{l=1}^{p} q_\mu^l \tag{2.1}$$

$$x_\mu = c_\mu^l - \bar{x}_\mu \quad \text{or} \quad x_\mu = q_\mu^l - \bar{x}_\mu \tag{2.2}$$

The covariance matrix, M, its eigenvectors, V, and the diagonal matrix of eigenvalues, $\mathbf{\Sigma}$, are given by

$$M_{\mu\nu} = \frac{1}{p}\sum_{k=1}^{p} x_\mu^k x_\nu^k = \frac{1}{p}\mathbf{X}^T\mathbf{X}, \qquad \mathbf{V}^T\mathbf{M}\mathbf{V} = \mathbf{\Sigma} \tag{2.3}$$

where $X_{k\mu} = x_\mu^k$. The eigenvectors of the covariance matrix are the principal components, and their associated eigenvalues are the variances of the path along the corre-

sponding coordinates. Alternatively, the eigenvectors of $\mathbf{M}$ can be obtained without computing $\frac{1}{p}\mathbf{X}^T\mathbf{X}$ by singular value decomposition of the matrix $\frac{1}{\sqrt{p}}\mathbf{X}$

$$\frac{1}{\sqrt{p}}\mathbf{U}^T\mathbf{X}\mathbf{V} = \mathbf{S}, \quad \mathbf{S}^T\mathbf{S} = \frac{1}{p}\left(\mathbf{U}^T\mathbf{X}\mathbf{V}\right)^T\left(\mathbf{U}^T\mathbf{X}\mathbf{V}\right) = \mathbf{V}^T\mathbf{M}\mathbf{V} = \mathbf{\Sigma} \tag{2.4}$$

where S is an appropriately dimensioned matrix containing the singular values on the diagonal. The singular values correspond to the standard deviation along the associated coordinates in V.

The above procedure yields a new set of coordinate axes that can be used to represent the path

$$\mathbf{R} = \mathbf{X}\mathbf{V} \qquad r_\lambda^k = \sum_\mu^{ncrd} x_\mu^k V_{\mu\lambda} \tag{2.5}$$

where the $R_{k\lambda} = r_\lambda^k$ are the coordinates rotated onto the principal component axes. If the eigenvalues or singular values are arranged in descending order, the associated coordinates have the property

$$\sum_{k=1}^{p} \left|r_\lambda^k\right| > \sum_{k=1}^{p} \left|r_{\lambda+1}^k\right| \tag{2.6}$$

for all values of λ. Only the first $d \ll 3N$ of these coordinates have chemically significant values for any of the images along the path. The path can be reconstructed using only these d coordinates without a loss of chemical accuracy.

$$\tilde{x}_\mu^k = \bar{x}_\mu + \sum_{\lambda=1}^{d} r_\lambda^k V_{\lambda\mu} \tag{2.7}$$

To assess the accuracy of the reduction, the geometries can be reconstructed with different values of d and compared to the original path in the full coordinate space. The difference, in Cartesian coordinates, between the reconstructed path versus the original

path is used as an error metric. To carry out this comparison with paths reconstructed using redundant internal coordinates, the Cartesians are generated iteratively by fitting displacements in Cartesians to displacements in redundant internals[11]. This process begins by using the Cartesians corresponding to the reactants as a starting point. Each image along the path is fit, iterating until the RMS change in the Cartesian coordinates is less than 10^{-4} bohr. The Cartesian coordinates of the reconstructed paths can be used to generate an energy profile for the reaction. Comparison to the energy profile from the original reaction path can then be used to assess the convergence of the energy as additional coordinates are included in the principal component expansion.

The principal components produced by the above approach will naturally depend on the number of images used in the analysis. A subset of the original path can be constructed by taking the transition state plus every nth image downhill in each direction. The PCA method can be applied to these paths to obtain a new set of $\bar{\mathbf{x}}$, $\mathbf{X}$ and $\mathbf{V}$. These new average coordinate values and principal components can then be used to reconstruct the path, using all of the images, and compare it to the original path. This provides a relationship between the number of images used in the PCA, the number of components used in the reconstruction of the path, and the reconstruction error.

A test suite of representative reactions was chosen to study using this coordinate reduction procedure. The transition state for the reaction was found, and a steepest descent reaction path was calculated using the Hessian based predictor corrector integration method[35] in the Gaussian 09 suite of programs[36]. The step size was set to 0.05Å and up to 200 points were computed in each direction. Mathematica[37] was used to analyze the data and calculate the reduced coordinates. For redundant internal coordinates, a single optimization step was taken from the reactant and product geometries found by the reaction path calculation in order to generate a set of internal

coordinate definitions. The union of these two sets was used to analyze the reaction in redundant internal coordinates.

2.3 Discussion and Results

The convergence behavior of the coordinate reduction procedure was tested with the ene reaction. This reaction is a condensation of a molecule with a double bond and an allylic hydrogen and a molecule containing a multiple bond (Figure 2.1). This system has only a few atoms, but the reaction requires is a concerted rearrangement of two double bonds and a hydrogen transfer. In the simplest ene reaction, propene and ethylene reacts to produce 1-pentene.

The effect of the PCA on the coordinates representing the path is illustrated in Figure 2.3. The Cartesian coordinates with the first, second, third and tenth largest variance for the ene reaction are plotted in Figure 2.3a. Note that each coordinate varies around its average value by ca. ± 1Å and that the tenth coordinate varies roughly as much as the first three. Figure 1b shows the first, second, third and tenth Cartesian principal components. The first principal component is almost linear and has a range of nearly 3Å. The tenth principal component is nearly constant when compared to the first three components.

Figure 2.4 shows the energy profiles corresponding to the paths defined by the first d principal components for $d = 1$ through 4 for Cartesian and redundant internal coordinates. With both coordinate systems, the region near the transition state is first to converge. Because the transition state occurs in roughly the middle of the reaction, the mean value of the coordinates is a better approximation to the transition state than it is to either of the minima. The redundant internal coordinates converge more quickly to the true energy profile because they are a more natural representation of chemical motion, particularly when the principal motion involves the rotation of a

dihedral angle. In either coordinate system, though, only a few degrees of freedom are needed to converge to the energy. Figure 2.5 shows the energy convergence for the ene reaction as a function of the number of principal components used to reconstruct the path. The dashed horizontal line marks the chemical accuracy threshold of 1 kcal/mol.

In general, fewer principal components are needed to achieve chemical accuracy in the energy than in the coordinates. Figure 2.6 demonstrates the convergence behavior of Cartesian and internal coordinates relative to the number of principal components needed to achieve chemical accuracy. The dashed horizontal line marks the chemical accuracy threshold of 0.005 Å. With this level of accuracy, the scalar curvature along the reaction path in the reduced coordinates is essentially the same as for the full set of coordinates.

The effect of using fewer images to represent the path is shown in Figure 2.7, where the error in the coordinates is plotted as a function of the number of images and the number of Cartesian principal components used to reconstruct the path. There is very little difference in the accuracy of the path versus the number of components used when more than 20 images are used, out of a total of 178 in the original path. When fewer images are used, the error increases regardless of the number of principal components used in the reconstruction of the path. The inset in the figure shows the mean absolute error in the average coordinates, $\bar{x}_\mu$, as a function of the number of images in the path, and this error does indeed reflect the error in the contour plot. This systematic error is a consequence of the mean-centered data used in the PCA method. For the most reliable results, the images should be spaced to provide an adequate representation of the average value of the coordinates. Furthermore, the number of images should be at least two times the number of principal components needed to achieve chemical accuracy.

Six representative test reactions were chosen to study the properties of the coordinate reduction method (see Figure2.1):

1. The ene reaction of ethylene with propene.
2. The Diels-Alder reaction of ethylene with cyclobutadiene to form norbornene.
3. Hydrolysis of acetamide by water with an additional water molecule.
4. The ring closing step of a Robbinson annulation.[38]
5. The trans-gauche rotational barrier of 1,2-dichlorethane.
6. The isomerization for alanine dipeptide from the 7_{ax} minima to the α_L minima. [39]

The transition state optimization and IRC calculation were carried out at the HF/3-21G level of theory for each reaction except numbers 2 and 4, which used instead the PM6 semiempirical method[40]. These examples were analyzed for the convergence to chemical accuracy of both the energy and the coordinates. The first 4 reactions have cyclic transition states which involve the concerted motion of many atomic centers. Reactions 5 and 6 are driven primarily by the rotation about 1 and 2 dihedral angles, respectively. Dihedral rotations are difficult to model using Cartesian coordinates, but are one of the primary coordinate types used to represent molecules when using internal coordinates.

Table 2.1 summarizes the energy convergence behavior for the reactions in the above test set, showing the number of principal components needed to achieve chemical accuracy in mean absolute error and max absolute error in the energy for every image along the path. In every case, the number of principal components is much smaller than the total number of coordinates. Redundant internal coordinates do a bit better

than the Cartesians, especially for reactions dominated by torsions. The convergence behavior for the coordinates is examined in Table 2.2. The redundant internals perform a little bit worse in this case; but again, there is a large difference between the total number of coordinates and the number in the reduced set for each reaction studied.

In addition to the above test reactions, three examples from the recent literature were also analyzed for their convergence to chemical accuracy of the coordinates. (See Scheme 2.2):

1. The tautomerization of 2-pyridone assisted by two water molecules.[41]

2. NO insertion into the Co-CH_3 bond of $CpCo(CH_3)(NO)$.[42]

3. Matrix metalloproteinase 2 (MMP2) inhibition by (S) SB-3CT.[43]

The first reaction involves two water molecules in a concerted proton shuttle to convert 2-pyridone to 2-hydroxypyridine. It was calculated using MP2 with a large basis set (6-311+G(d,p)) and illustrates that the coordinate reduction methodology works equally well when more accurate levels of theory are employed. The second reaction is an example inorganic reaction involving NO migratory insertion into a Co-CH_3 sigma bond, and was computed with the B3PW91 functional[44, 45, 46, 47] and the 6-311G(d) basis set. The third reaction is a large scale QM/MM study of the mechanism of SB-3CT reacting in the active site of MMP2 and inhibiting the enzyme. The reaction path involves a concerted proton abstraction, opening of a three-membered ring and tight binding of a thiolate with the zinc ion in the active site. This example illustrates that the reaction path for even a very large system can depend on only a few coordinates.

Since the energy converges before the coordinates in every example from the test set, only the coordinates were analyzed in the three examples from the literature. These reactions show behavior similar to those of the test set, where fewer than 15 principal components are needed even though many more coordinates define the full system.

Even in the extreme case of the QM/MM reaction path where there are thousands of atoms involved, only 6 Cartesian principal components are necessary to achieve chemical accuracy. For large systems, the maximum error is a better measure of the number of coordinates needed, since the mean absolute error is an average over a large number of nearly stationary atoms.

2.4 Conclusions

Principal component analysis is an effective technique to significantly reduce the number of coordinates necessary to accurately reproduce the geometric changes along a reaction path for a chemical reaction. For the representative reactions studied here, the reduction often exceeded 90%. Such a dramatic reduction may be useful in reaction path optimization, provided a method for updating the reduced space during the optimization is developed. Additionally, coordinate reduction may also be advantageous in reducing the cost of constructing and evaluating high accuracy interpolated potential energy surfaces for studying reaction dynamics.

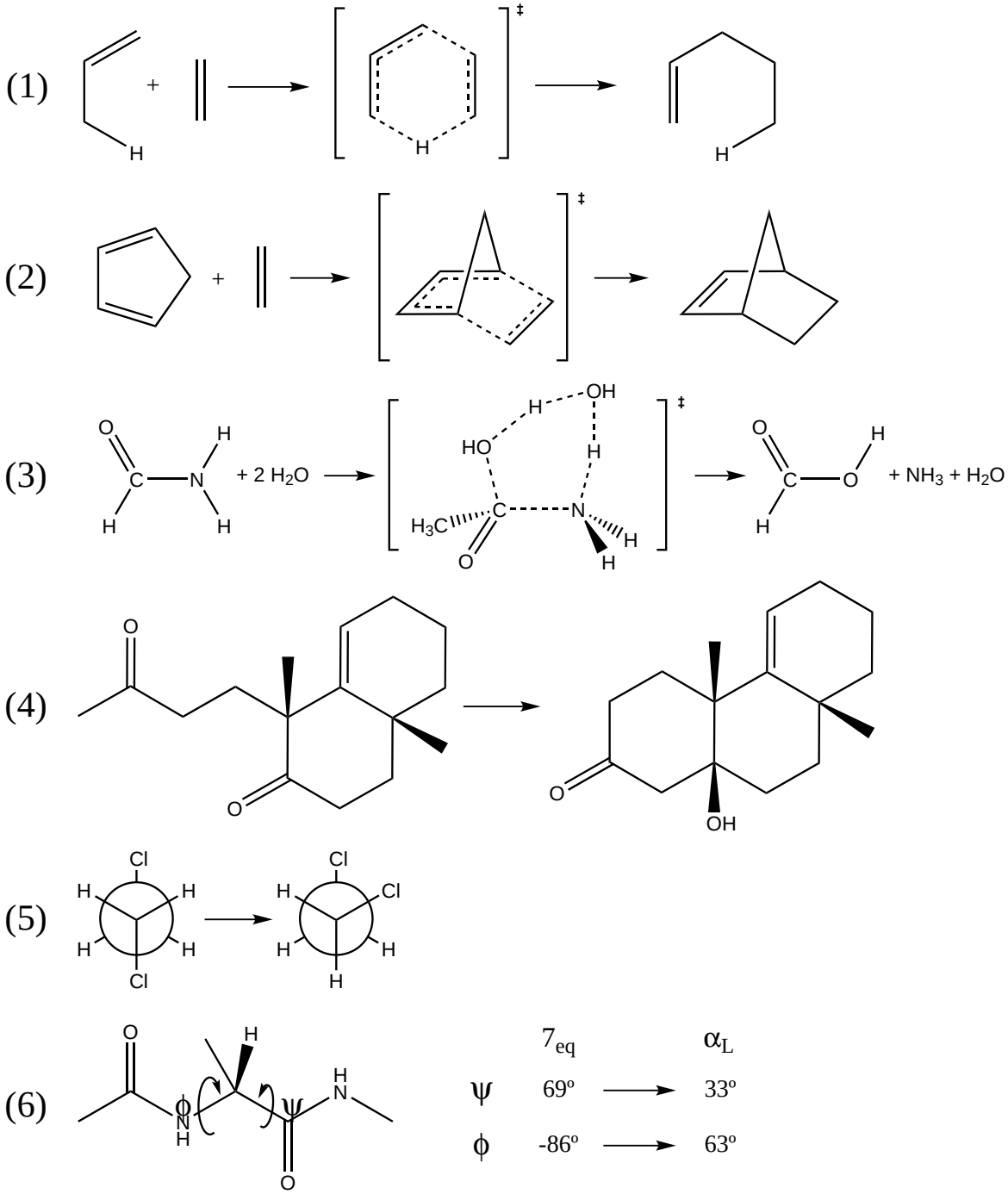


Figure 2.1: Representative test reactions

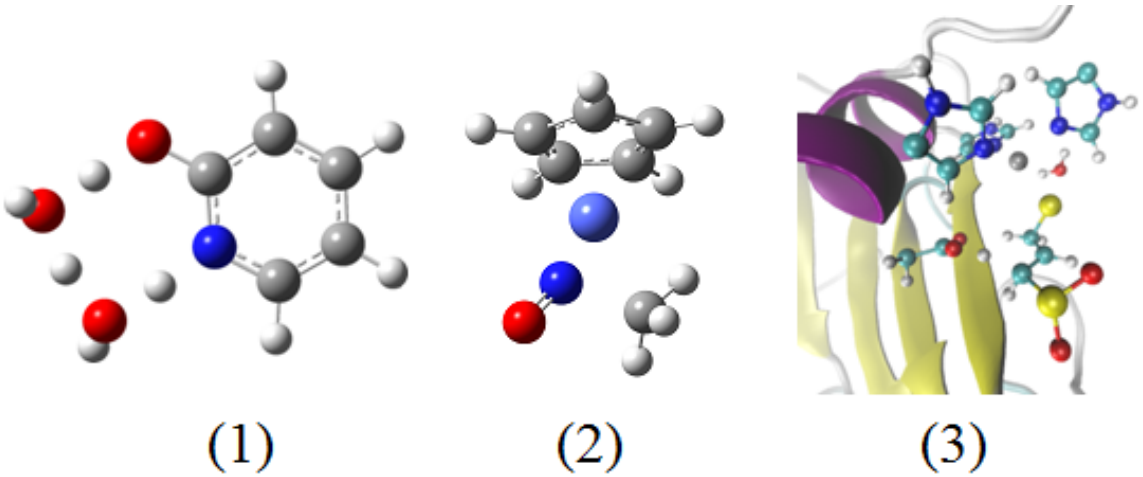


Figure 2.2: Additional test reactions

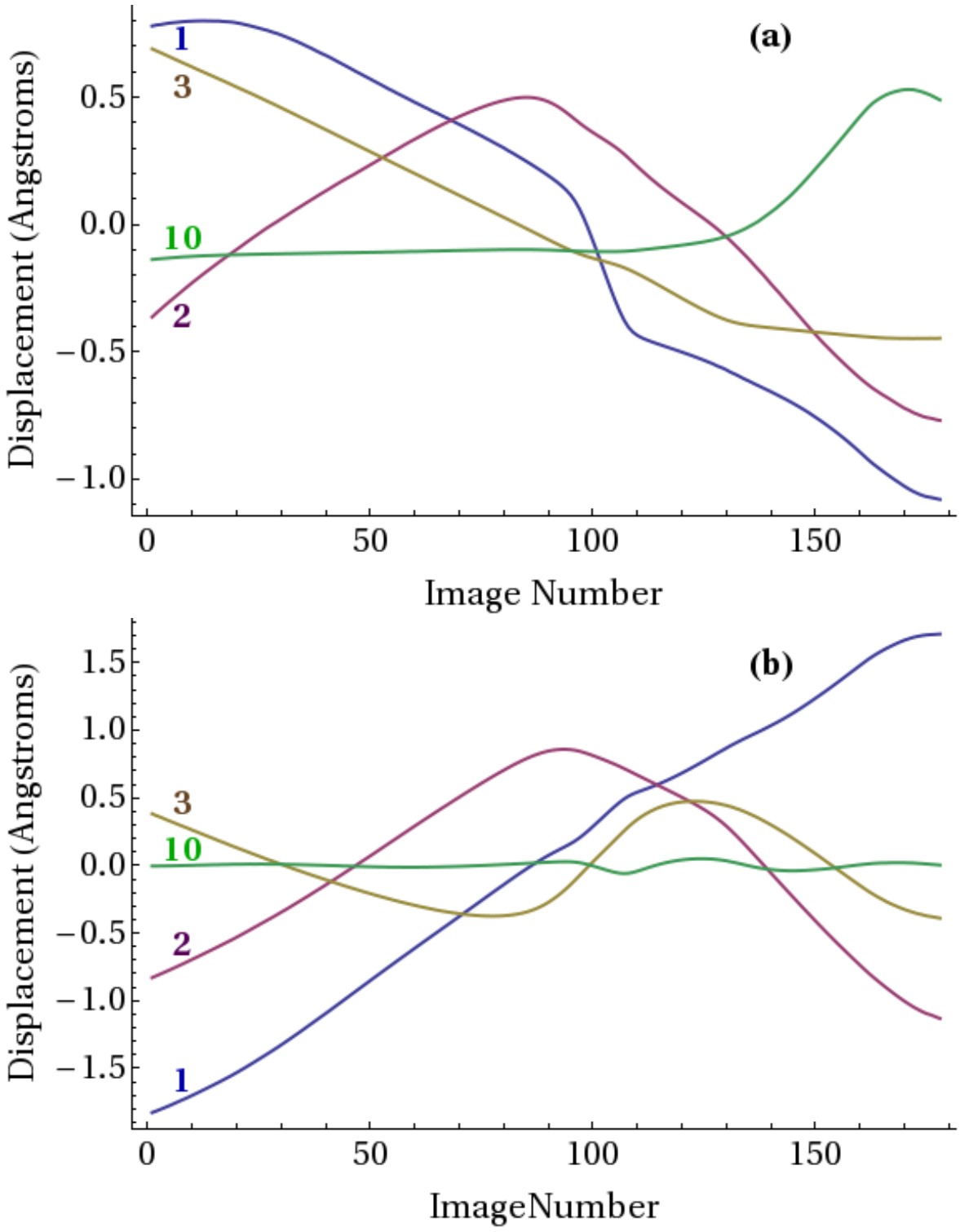


Figure 2.3: Reaction path for the ene reaction. **a** The three Cartesian coordinates with the largest variance and the Cartesian coordinate with the 10th largest variance. **b** The three Cartesian principal components with the largest variance and the Cartesian principal component with the 10th largest variance.

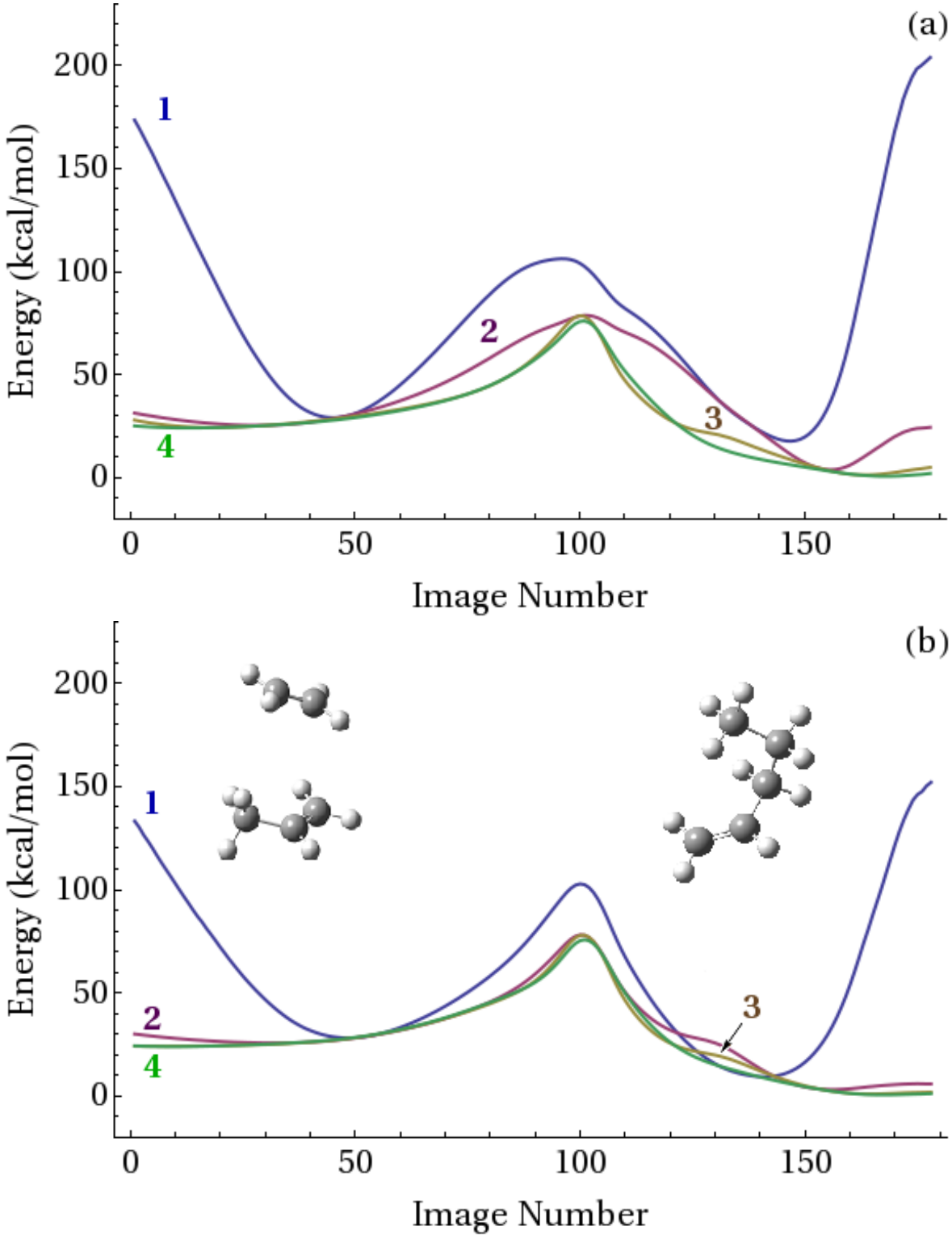


Figure 2.4: Energy along the reaction path for the ene reaction using the first d principal components to represent the reaction path, $d = 1 - 4$. **a** Cartesian coordinates. **b** Redundant internal coordinates

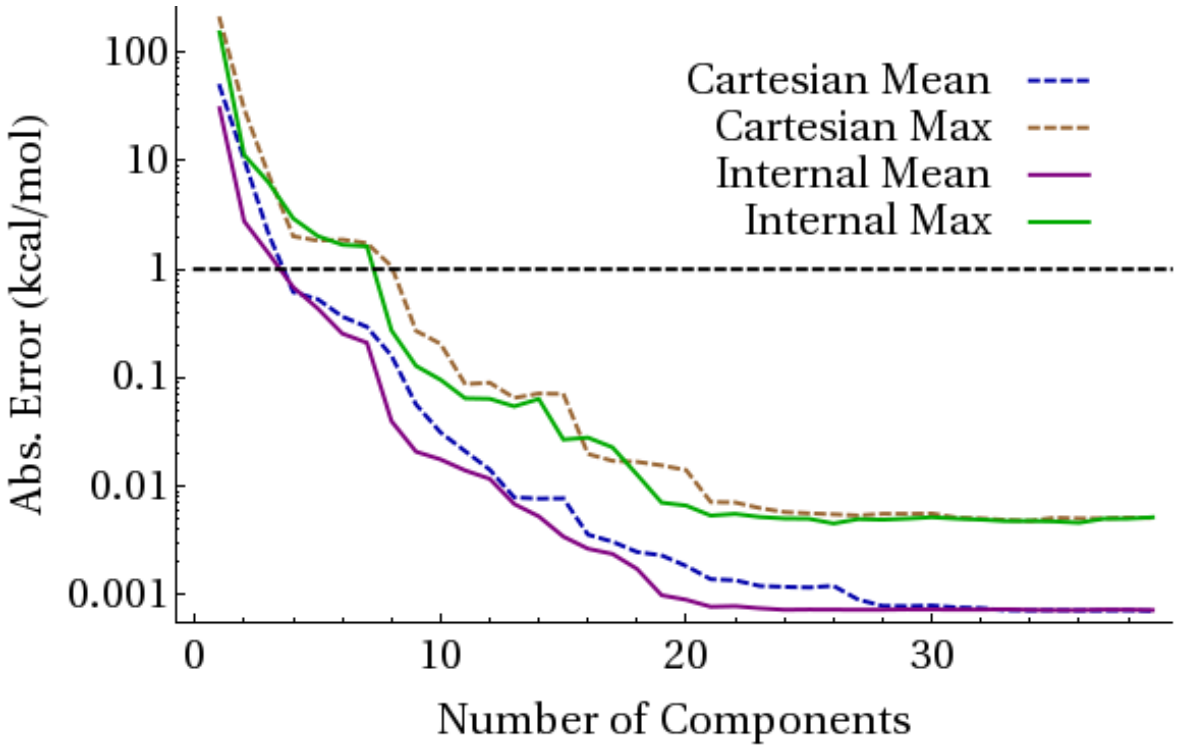


Figure 2.5: Mean absolute error and maximum absolute error in the energy along the path for the ene reaction as a function of the number of principal components, the dashed horizontal line indicates chemical accuracy of 1 kcal/mol

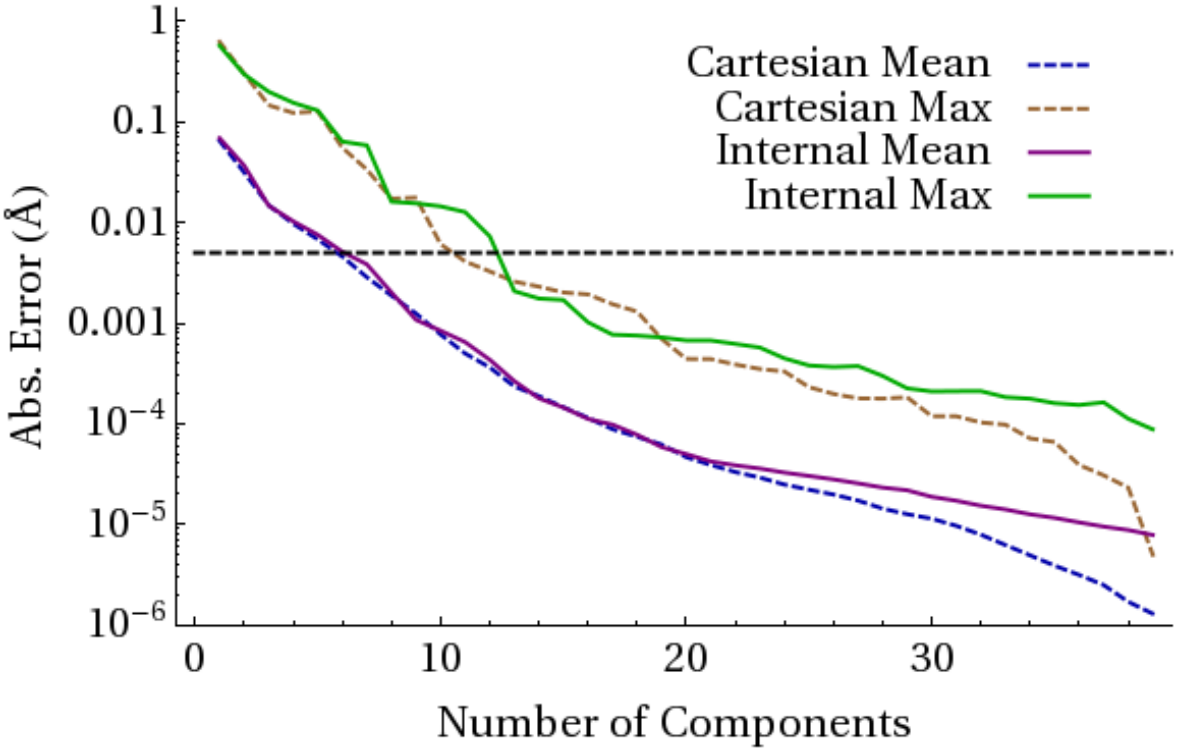


Figure 2.6: Mean absolute error and maximum absolute error in the Cartesian coordinates along the path for the ene reaction as a function of the number of principal components. The dashed horizontal line indicates chemical accuracy of 0.005 angstroms

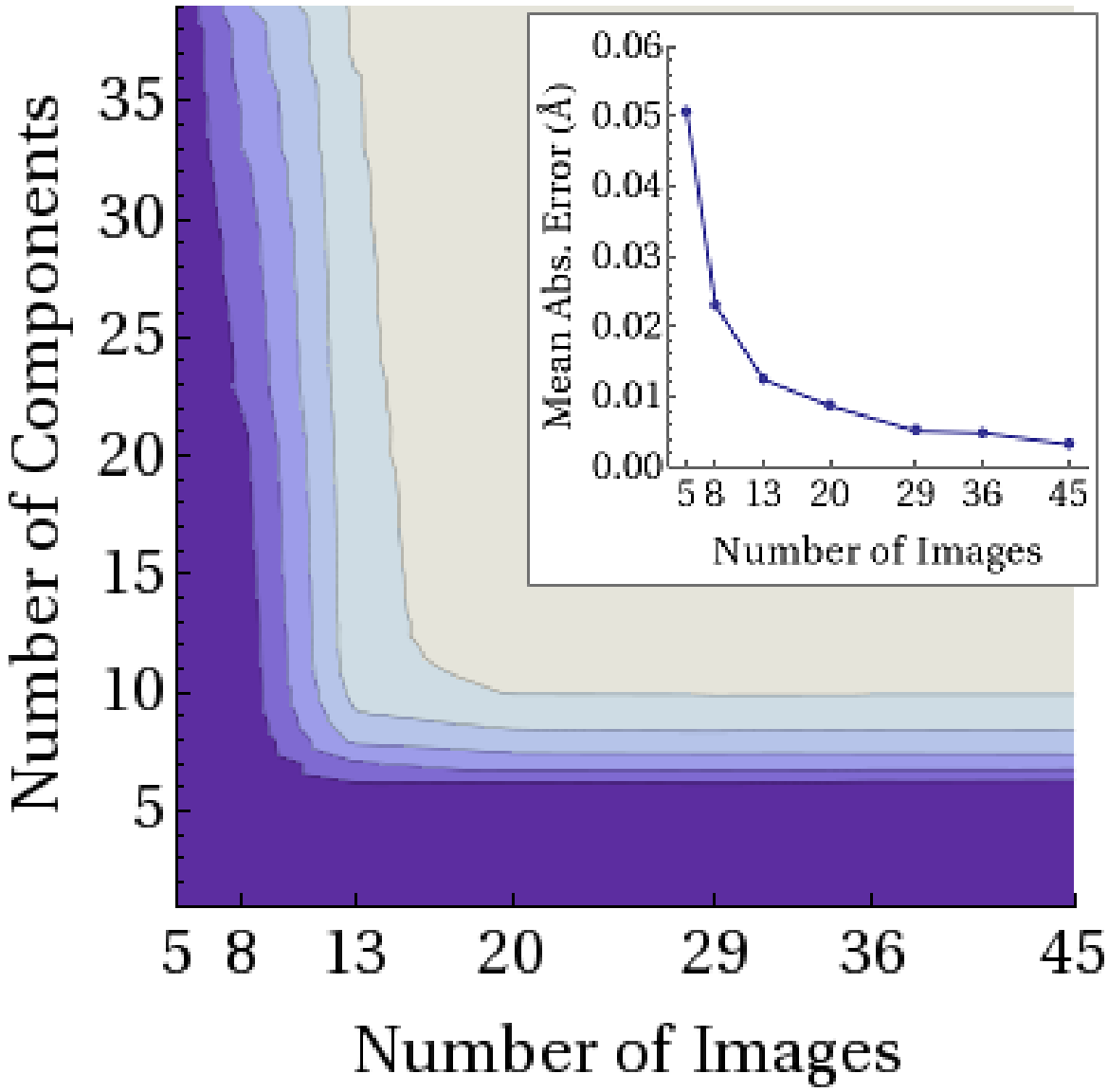


Figure 2.7: Contour plot of the mean absolute error in the coordinates along the path for the ene reaction as a function of the number of principal components and the number of images used to generate the principal components, with the darkest region corresponding to an error of 0.005 angstroms. The inset shows the mean absolute error in the computation of the average coordinates versus the number of images in the subset path

Reaction	Cartesian		
	# Total	# PCA mean	# PCA max
Ene	45	4	9
Diels-Alder	57	3	5
Amide hydrolysis	45	5	8
Robinson annulation	126	5	8
Dichloroethane	24	2	2
Alanine dipeptide	66	3	5

Reaction	Redundant Internal Coordinates		
	# Total	# PCA mean	# PCA max
Ene	79	4	8
Diels-Alder	244	3	5
Amide hydrolysis	112	5	7
Robinson annulation	283	4	7
Dichloroethane	28	1	1
Alanine dipeptide	98	1	2

Table 2.1: Comparison of the number of principal components needed to achieve chemical accuracy (<1 kcal/mol) in the mean and maximum absolute error in the energy of the reduced path

Reaction	Cartesian		
	# Total	# PCA mean	# PCA max
Ene	45	6	11
Diels-Alder	57	3	8
Amide hydrolysis	45	7	14
Robinson annulation	126	5	12
Dichloroethane	24	2	4
Alanine dipeptide	66	4	8
Pyridone + 2 H_2O	54	4	9
CpCo(NO)(CH_3)	51	3	5
MMP2	7932	1	6

Reaction	Redundant Internal Coordinates		
	# Total	# PCA mean	# PCA max
Ene	79	7	13
Diels-Alder	244	3	7
Amide hydrolysis	112	7	15
Robinson annulation	283	6	14
Dichloroethane	28	1	3
Alanine dipeptide	98	5	8

Table 2.2: Comparison of the number of principal components needed to achieve chemical accuracy (0.005 angstroms) in the mean and maximum absolute error in the Cartesian coordinates of the reduced path

CHAPTER 3

USING BONDING TO GUIDE TRANSITION STATE OPTIMIZATION

3.1 Introduction

The optimization of equilibrium geometries and transition states is an important step in the computational study of chemical reactions (see references[48, 2] for overviews of current methods). In general, geometry optimization begins by computing the energy and forces acting on a guess structure and then explores a local approximation to the potential energy surface with the goal of locating a structure where the forces are zero (the forces are the negative of the potential energy gradient). This is accomplished by an iterative process, where each successive optimization step produces a geometry closer to the solution, and new energies and forces are computed to update the local approximation. The commonly used Newton-Raphson type optimization methods utilize information about the second derivatives of the potential energy surface (PES), the Hessian matrix, in order to explore the surface with greater confidence than using only the forces alone in order to reduce the number of optimization steps required. The full Hessian matrix can be computed analytically at a significantly higher computational cost than required to produce an energy and gradient. Alternatively, a few of the eigenvectors and eigenvalues may be computed analytically or numerically at a reduced cost[49, 50, 51]. For quasi-Newton optimizations seeking a minimum energy structure it is usually sufficient to begin with a diagonal, positive definite estimate based upon empirically determined values[29, 52, 53]. Estimating the Hessian for tran-

sition state optimization by the quasi-Newton method is more difficult, as the Hessian matrix must have exactly one negative eigenvalue.

Quasi-Newton methods begin with an analytical or estimated Hessian matrix, and then update it using gradients calculated during the course of the optimization using one of a variety of different Hessian update schemes (for example, see references[27, 22, 23, 24, 25, 26, 28]). When using a (quasi-)Newton Raphson method to locate transition states, the Hessian matrix must have exactly one negative eigenvalue. The eigenvector corresponding to this eigenvalue is also called the transition vector, and the energy along this vector is a maximum at the transition state. If the initial structure to be optimized is sufficiently close to the actual transition state geometry, the exact Hessian matrix computed by analytical or numerical means will have only one negative eigenvalue and a Hessian update formula that allows for negative eigenvalues[26, 27, 28] may be used to converge to the transition state geometry. However, it is often difficult to produce such a good guess geometry, and other approaches must be used, such as selecting which eigenvector to follow uphill and then correcting the Hessian before computing the step[54], or utilizing a reduced potential surface to restrict the transition state search to the dominant coordinates in the reaction[55, 56, 57].

If appropriate estimates of the reactant and product structure are used, the transition vector can be estimated by the tangent to a simple interpolated pathway[32], and an empirical estimate of the Hessian may be used after some initial steps are taken along the approximate pathway in order to correct the curvature[58]. However, these estimates tend to be a good approximation to the reaction path only when the structures corresponding to the reactant and product in the interpolation lie close to the barrier on the potential energy surface. In practice, when fully optimized reactant and product geometries are used, considerable re-orientation of the geometry may be required before the reaction can occur. Methods such as the dimer method[59] or one

of many reaction path optimization methods[4, 5, 6, 60, 61] may be used to explore the surface in order to locate a region that is likely to contain the transition state. These produce good initial guess structures for a quasi-Newton transition state optimization and in some cases may also be used to generate a suitable approximate or updated Hessian[62, 63] that rapidly accelerates convergence to the transition state geometry at the cost of *many* gradient calculations prior to the beginning of the transition state search.

Previous work[64] demonstrated that a reaction pathway may be accurately described by a small number of composite coordinates constructed as linear combinations of primitive Cartesian or internal coordinates. When the transition state represents the energy barrier for a simple conformational change in geometry, a combination of many bending, stretching and torsional coordinates may contribute significantly to the motion along the path near the transition state, and it can be difficult to determine which of these coordinates are necessary to represent the transition vector from the reactant and product geometries alone. However, for a chemical reaction, the desired transition state necessarily involves the forming and/or breaking of bonds. In the following work, we attempt to use knowledge about the bonding in reactants and products to produce a guess structure for the geometry at the transition state. Additionally, this bonding information is used to decrease the need for the computation of an analytical Hessian, and to improve the selection of the transition vector during the optimization in order to reduce the number of gradient calculations needed to locate a transition state. In order to compare the performance of our proposed methods, 20 reactions were selected from a variety of published[58, 65, 12, 66] and unpublished libraries for use as a test set (see Figure 3.1 for details). This test set contains characteristic examples of many types of organic reactions including insertions, additions, eliminations, hydrolysis, ring openings, substitutions, cycloadditions and rearrangements. Reactions

containing transition metals will be examined in future work.

3.2 Methods

3.2.1 Choice of Coordinate System

The first step in any geometry optimization algorithm is the selection of a coordinate system to describe the changes in the geometry. While the Cartesian coordinates of the nuclear centers are an intuitive and straightforward set of coordinates to use, a redundant set of internal coordinates[11] (bond stretches, angle bends, and dihedral torsions) are a more natural basis for describing the potential energy landscape for a chemical system. Such coordinate systems reduce the coupling between the various coordinates, resulting in a much more diagonally dominant Hessian matrix than typically observed in a Cartesian representation. This is especially important for optimization methods that rely on Hessian updating, and hence redundant internal coordinate systems have enjoyed a wide-spread use in geometry optimization.

Any number of different coordinate sets may be constructed from the possible combinations of bond stretches, angles and dihedrals, so long as enough coordinates are included to completely describe the internal degrees of freedom of the system. The standard approach to define a set of redundant internal coordinates for optimization is to begin by determining the bonding skeleton of the structure. This can be input to the optimizer by the user, or it can be generated automatically by computing all of the distances between atoms in the molecule, and considering any two atoms bonded if they are roughly as close as, or closer than, a reference single bond length for those two atoms. Once the connectivity of the structure is known, the coordinate system may be constructed by including the bond stretches between any bound atoms, the bends between any pairs of stretches that share an atom, and dihedral angles between bends that share a bond. Once this set is defined, extra care must be taken to be sure to

include any coordinates necessary to describe the bending of a near-linear angle, the motion into/out of a plane for atoms in a nearly planar configuration, as well as any coordinates necessary to describe the relative motion of two unconnected fragments in the geometry.

This approach is complicated when selecting a coordinate set for optimization of the transition state for a chemical reaction. In this case, the bond lengths corresponding to bonds breaking would be much longer in the product structure, and bond lengths corresponding to bonds forming are going to be significantly longer in the reactants. If the reactant or product structures are multi-molecular, additional "virtual" bonds must be added, usually joining the pair of atoms on two molecules with the shortest distance between them. This will result in a coordinate set that includes coordinates to describe the relative motion between fragments, but the added virtual bond will not necessarily be a bond involved in the reaction when significant reorientations of the molecules are required to move from the minimum structure to the region of the potential energy surface where the transition state lies. Including such coordinates can cause problems in an optimization and should be avoided. Likewise, special coordinates such as those needed to describe the bending motions of angles that are nearly linear, might be appropriate for describing the reactant and/or the product even when the corresponding angle is non-linear at the transition structure, and the inclusion of such coordinates can also frustrate an optimizer. For these reasons, taking the union of coordinates defined at the reactants and products may not be the best approach for a transition state optimization.

The approach described in this chapter, instead, merges the bonding skeletons from the reactant and product structures, and then uses the merged skeleton to define the coordinate set. For the purpose of interpolating an initial geometry, an initial coordinate set may be defined by adding or removing the necessary coordinates to

describe the linear or planar structures in the reactants and the products. Once the initial guess for the transition state is generated, the coordinate set is redefined using the geometry of the guess, so that only the linear and planar coordinates necessary to describe the interpolated structure are included. Defining the coordinates in this way, the number of reactions that fail to produce an initial geometry by linear interpolation of the reactants and products is reduced from 6 to 2 for the 20 reactions in our test suite. To further improve the reliability of interpolating reactant and product structures to produce an approximate transition state geometry, a new approach must be introduced.

3.2.2 Bond Order Interpolation with Relaxation

To find the transition state corresponding to a reaction where the minimized structures of the reactants and products are already known, the typical approach is to begin by interpolating between the two structures to generate a better guess for the transition state optimization. This is usually done by a simple linear interpolation, often in distance matrix or internal coordinates, to the midpoint of the line connecting the reactants and products. This approach includes an explicit dependence on the reactant and product geometries, which can result in wildly different guess structures when different, equally valid, geometries are input. A simple example would be in the case of a bimolecular reaction, the interpolated midpoint is highly dependent upon the distance between the two reactants. Typically, only a few of the total set of coordinates that describe the system are involved with the reaction, and it is better to limit the interpolation to that reduced set. Intuitively, the bonds that are breaking or forming over the course of the reaction must belong to that set, and it should be sufficient to interpolate along only those coordinates.

In order to use the bonding information to interpolate a guess of the structure at the transition state, it is necessary to have a convenient description of bonding. By using

the concept of bond order, a relationship between a bond's length and approximate strength can be described. Pauling[67] suggested using the following equation

$$o\left(r\right) = e^{(r_0 - r)/a} \tag{3.1}$$

$$r\left(o\right) = r_0 - a\ln\left(o\right) \tag{3.2}$$

where o is the order, r_0 is a reference single bond length for the atoms involved, and a is a positive constant determined by fitting to known experimental or computational data. With $a = 0.3\mathring{A}$, this equation agrees well with typical values for double and triple bonds. Unfortunately, bond order initially decays quite rapidly when r is longer than a single bond length, and there is a smooth decrease in the bond order out to infinity. Using equation (3.2), a bond of order 0.5 would be roughly 114% the length of the reference single bond length, while the associated bond length will increase slowly to infinity as the order approaches 0. These are characteristics that are undesirable for interpolation of bonds in transition state geometries.

Instead, the present work will use the following approximate bond order for interpolation of the bond order

$$\begin{aligned} \widetilde{o}(r) &= \max\left[0, \frac{\left(b\left(r_0/r\right) - 1\right)}{\left(b - 1\right)}\right] \\ r\left(\widetilde{o}\right) &= \frac{b}{\left(\left(b - 1\right)\widetilde{o} + 1\right)} r_0, \;\; \widetilde{o} \geq 0 \end{aligned} \tag{3.3}$$

While this form is not as accurate as the Pauling bond order for bond orders greater than one, it decays much more quickly as the bond length increases, and will assign all bond lengths greater than a particular value (determined by selection of b) an order of

zero. A typical bond length at a transition state for a bond being broken or formed is between roughly 120-150% the length of the corresponding standard bond length, and it is usually safe to assume that when the distance between two atoms is greater than twice the standard bond length they are not bonded. Using a value of 2 for b in the equations above results in a bond length 133% of the reference at bond order 0.5, and a bond order of 0 for bond lengths greater than or equal to 200% the reference length. Table 3.1 lists the mean, maximum and minimum bond order of the bonds being broken/formed for the geometries in the test set computed at the PM6 level of the theory, and demonstrates that a bond order of roughly 0.5 produces reasonable bond lengths for a wide variety of reactions.

Using this definition of bond order, the linear interpolation can be restricted to the bond order of only the bonds being broken or formed during the reaction. A direct, least-squares fit to the change in the bonding coordinates that is necessary to achieve the interpolated structure may cause other parts of the geometry to change in unfavored ways, and bring atoms too close to one another. In order to maintain reasonable values for the remaining coordinates, a relaxed scan that minimizes an inexpensive approximate potential energy orthogonal to the set of bonds is used.

Molecular mechanics potentials like the Universal Force Field (UFF)[68] are among the least expensive approximations to the energy of a chemical system, and should be sufficient for the purpose of relaxing the coordinates orthogonal to the bonds being broken or formed. The UFF force field is simple and parameterized for a large number of atom types, and when the combined connectivity skeleton from the reactants and products is used to select the atom types, it is generally sufficient to produce bonds and angles that avoid very unfavorable regions of the actual potential. Alternatively, semi-empirical electronic structure methods such as PM6[40] may produce structures that are better suited for optimization of transition states using Hartree-Fock or density

functional theory. To avoid significantly impacting the total cost of the optimization, the low level method used during interpolation and relaxation should be at least an order of magnitude faster than the electronic structure method used for the transition state optimization. For different classes of reactions other than the representative organic reactions discussed in the present work, such as those involving transition metals, the selection of the appropriate low level method for this interpolation process may require further investigation.

The proposed interpolation algorithm is as follows:

1. Determine which atom pairs are bound for both the reactant and product. Compare the lists of bonds to determine which bonds are breaking/forming, and combine the lists of bonds to construct a redundant coordinate set.

2. Compute the bond orders for the breaking/forming set in both reactants and products, and compute a set of goal bond lengths for these bonds from the average value of the bond order.

3. Clean up the reactant and product structures by fitting the displacements to bring any bond lengths longer than 2 times the reference length down to 2 times the reference length, and then minimize the low level energy orthogonal to the breaking/forming set.

4. Beginning with the reactant structure, take a series of steps no greater than 0.5 bohr along the displacement to the goal lengths. Following each step, the low level energy is minimized by updating the coordinates orthogonal to the breaking/forming set.

5. Repeat 4, beginning with the product structure.

6. Select the guess with the lower energy, evaluated by the low level method, as the initial geometry to optimize to a transition state on the actual potential.

When force field methods are used for interpolation and relaxation, some care is needed for defining the atom types for atoms that undergo a change in bonding. The atom types for these atoms are selected based upon the number of connected groups in the combined connectivity skeleton if such an atom type is defined, otherwise the atom type is set to the atom type with the largest number of connected groups for that atom.

When this algorithm is applied to interpolating an initial guess structure for each of the 20 reactions in the test set, a structure with reasonable bond lengths and angles is produced even when beginning with fully minimized reactant and product structures. In some cases, however, the optimization of the structures either fails to converge or converges to a transition state for an alternative reaction, particularly when Hessian updating is employed. In order to improve the reliability of the transition state search, new methods for selecting the transition vector, computing the optimization step, and estimating the Hessian for a transition state optimization were developed.

3.2.3 Using Connectivity Change to Approximate the Transition Vector

Once an approximate structure has been obtained, a transition state optimization needs to be carried out to find the structure that maximizes energy along the transition vector while minimizing the energy along all other coordinates. The Newton-Raphson method can accomplish this by employing a shifted Hessian to compute the displacement towards the stationary point with the desired curvature

$$\Delta \mathbf{q} = -\left(\mathbf{H} + \mathbf{S}\right)^{-1} \mathbf{g} \tag{3.4}$$

where $\mathbf{g}$ and $\mathbf{H}$ are the gradient and Hessian of the PES, and $\mathbf{S}$ is a matrix that is typically constructed to control the direction and the size of the displacement step. For a transition state optimization, the shifted Hessian must have exactly one negative eigenvalue. The rational function optimization (RFO) method defines $\mathbf{S}$ as a scaled identity matrix (i.e. $\mathbf{S} = -\boldsymbol{\lambda} \boldsymbol{I}$), where the scaling factor is the predicted energy change of the step on a local rational function approximation to the PES[54]. This scaling factor is computed straightforwardly as an eigenvalue of the augmented Hessian

$$\mathbf{H}_{\boldsymbol{aug}} = \begin{pmatrix} \mathbf{H} & \mathbf{g} \\ \mathbf{g^T} & \mathbf{0} \end{pmatrix} \tag{3.5}$$

For a geometry defined by n variables, the augmented Hessian has n+1 eigenvalues, and using the nth most negative eigenvalue results in a shifted Hessian with n-1 negative eigenvalues.

While the RFO method can compute a single correction such that the shifted Hessian has the correct number of negative eigenvalues, this will always result in following the eigenvector of the Hessian with the most negative eigenvalue uphill. The Partitioned RFO (pRFO) method selects a specific eigenvector to follow uphill, which may not necessarily be the one corresponding to the most negative eigenvalue, and defines $\mathbf{S}$ as the following sum

$$\mathbf{S}_{\boldsymbol{pRFO}} = -\lambda_1 v_{TS} v_{TS}^{\mathrm{T}} - \lambda_2 \sum_{i}^{n-1} v_i v_i^{\mathrm{T}} \tag{3.6}$$

where v_{TS} is the eigenvector whose corresponding eigenvalue should be negative and λ_1 is the most positive eigenvalue of the augmented Hessian constructed in the eigenspace defined by v_{TS}, while the v_i are the remaining n-1 eigenvectors of the Hessian and λ_2 is the most negative eigenvalue of the augmented Hessian defined in that space.

The most difficult part of a transition state optimization that employs the pRFO method is the selection of the eigenvector along which the energy should be maximized. One existing strategy that has been used effectively in the past is to compare the eigenvectors of the Hessian to the tangent of a simple approximation to the reaction path. These approximate pathways are typically functions of the geometries corresponding to the reactant and product, with some also including a dependence on the geometry at the current step of the optimization. Three such functions[32, 58, 69] that have been used in the past are a linear path connecting reactants ($\mathbf{q}_R$) and products ($\mathbf{q}_P$), a quadratic path connecting $\mathbf{q}_R$ and $\mathbf{q}_P$ through the current estimate of the transition state ($\mathbf{q}_{TS}$), and an arc of a circle connecting $\mathbf{q}_R$ and $\mathbf{q}_P$ through $\mathbf{q}_{TS}$. The tangent for the linear (τ_{lin}), quadratic (τ_{quad}), and arc of circle (τ_{arc}) paths are given by

$$\tau_{lin} = \mathbf{q}_P - \mathbf{q}_R \tag{3.7}$$

$$t_{quad} = \frac{\tau_{lin}^{\mathrm{T}} \left(\mathbf{q}_{TS} - \mathbf{q}_R\right)}{\tau_{lin}^{\mathrm{T}} \tau_{lin}} \tag{3.8}$$

$$\tau_{quad} = \tau_{lin} + \frac{\left(1 - 2t_{quad}\right)\left(\left(\mathbf{q}_{TS} - \mathbf{q}_R\right) - t_{quad}\tau_{lin}\right)}{t_{quad} - t_{quad}^2} \tag{3.9}$$

$$\tau_{arc} = \frac{\left(\mathbf{q}_P - \mathbf{q}_{TS}\right)}{\left|\mathbf{q}_P - \mathbf{q}_{TS}\right|^2} - \frac{\left(\mathbf{q}_R - \mathbf{q}_{TS}\right)}{\left|\mathbf{q}_R - \mathbf{q}_{TS}\right|^2} \tag{3.10}$$

see Bell et al.[69] for the derivation of the quadratic vector, and Peng et al.[65] for the derivation of the arc vector. All coordinate values are expressed as redundant internal coordinates.

These approaches can produce an accurate estimate of the transition vector when the reaction is relatively simple, or the reactant-like and product-like structures supplied are located quite close to the transition state, but they tend to be less reliable when fully optimized structures are used, and the reaction path has multiple, distinct

regions with different curvature. When the transition state involves bond breaking and forming, bond stretch coordinates tend to be the dominant coordinates in the transition vector. As an alternative to the above approaches, a simple "delta bonding" (Δb) vector is proposed, whose elements are defined in the following way:

1. Stretches corresponding to bonds forming in the reaction have the value 1
2. Stretches corresponding to bonds being broken in the reaction have the value -1
3. All other coordinates have a value of 0

This vector is trivial to compute, and does not depend explicitly on the geometry at the reactant, product or transition state. Additionally, for the reactions in the test set, Δb has a large overlap with the transition vector while typically having a small overlap with all remaining eigenvectors of the Hessian computed at the transition state geometry. Table 3.2 compares the linear, quadratic, and arc of circle tangents to the Δb vector for the set of reactions described in section 3. The vectors, in redundant internal coordinates, are computed according to equations (3.7-3.10) using the transition state geometry, and then projected into the locally non-redundant space

$$\tau_{NR} = \mathbf{B}\mathbf{B}^{+}\tau \tag{3.11}$$

where B is the Wilson B-matrix, which contains the partial derivatives of the internal coordinates with respect to a change in the Cartesian coordinates $B_{ij} = (\partial q_i/\partial x_j)$. The Δb vector not only has a large overlap with the transition vector in the majority of the reactions, but it also tends to have a small overlap with the remaining coordinates.

3.2.4 Initial approximation of the Hessian

For minimizations, the simple valence force field Hessian tends to work well since it is a rough, but reasonable, approximation to the exact Hessian at a minimum. This

reduces the amount of work that the optimizer and update formulas must do in order to improve it. Additionally, when the initial geometry is poor and forces are large, the shift matrix $\mathbf{S}$ in equation (3.4) will tend to dominate, resulting in similar optimization steps regardless of whether the Hessian used is exact or approximate. For transition state optimizations, however, the direction of the step is considerably more sensitive to the accuracy of the Hessian, particularly in how the transition vector couples to the minimization modes when far from the solution. Since the initial few steps of transition state optimizations will be largely downhill, it would be beneficial to make use of the increased stability of the downhill RFO steps and BFGS updates to get as close as possible to the transition state when using approximate or updated Hessian information.

The method described earlier to interpolate an estimated transition state geometry along the bonds being broken or formed during the reaction utilizes the a less expensive method to approximate the PES for the purpose of relaxing the geometry along the remaining coordinates. While these methods generally do a good job at predicting bond lengths and angles, they may neglect or simplify many of the non-bonding interactions that are more correctly accounted for in Hartree-Fock or DFT calculations. The resulting forces may still be quite large in the displacements corresponding primarily to the non-bonded interactions. Relaxing these modes on the same PES that the transition state optimization will be carried out on prior to beginning the transition state search should help to reduce some of the risks associated with using approximated and updated Hessian information.

Since it would be difficult to attempt to define a set of coordinates corresponding only to the non-bonded interactions, the minimization phase is carried out by choosing a small number of coordinates to be frozen while there remaining are allowed to relax. The frozen set contains the bonds being broken or formed, and angles between any pairs

of these bonds that share an atom. If the convergence criterion on the minimization phase is too loose, the geometry may not change enough to make a meaningful impact on the transition state optimization. Likewise, convergence criteria that are too tight may result in an unnecessarily large number of optimization steps in the minimization phase with no improvement to the transition state optimization, particularly in any case where the actual transition vector is not entirely contained in the frozen set. Using the RMS of the gradient in the active set of coordinates as the sole convergence metric, a threshold of $3.2x10^{-4}$ hartree/bohr was determined to provide a good balance between the time spent in the minimization phase and the efficiency and reliability of the transition state optimization. This is comparable to the convergence threshold used for regular geometry optimization.

During the minimization phase, the empirical approximation to the Hessian will work as well as it does in a standard minimization, and use of the BFGS update formula should be sufficient to achieve convergence. Following the minimization phase, however, it is likely that the approximate Hessian will be positive definite, making it a poor choice to begin the transition state search. Given that Δb is a good approximation to the transition vector at the solution, the initial Hessian for the transition state optimization may be constructed from the updated, positive definite Hessian from the minimization phase $\left(\widetilde{\mathbf{H}}\right)$ by reversing the sign and adjusting the magnitude of the Rayleigh coefficient along the locally non-redundant projection of Δb. Using τ_{NR} from (3.11), the initial Hessian $\mathbf{H_0}$ for the transition state optimization is given by

$$\mathbf{H}_0 = \widetilde{\mathbf{H}} - 1.5\frac{\tau^{\mathrm{T}}\widetilde{\mathbf{H}}\tau}{\tau^{\mathrm{T}}\tau}\frac{\tau\tau^{\mathrm{T}}}{\tau^{\mathrm{T}}\tau} \tag{3.12}$$

This simple transformation produces an approximate Hessian with the correct curvature for a transition state search, and that guarantees that the eigenvector with a

negative eigenvalue will have a large overlap with Δb.

3.2.5 Divided RFO

The shifted Hessian computed with the pRFO shift matrix (3.6) guarantees that the computed step will be downhill in the minimization space, and uphill along the chosen transition vector. This is a consequence of the fact that the scale factor in the minimization space is more negative than the most negative eigenvalue in that space, and the scale factor along the transition vector is more positive than its corresponding eigenvalue. When the Hessian information is accurate, the step size computed by this approach tends to be reasonable even when the minimization space is not positive definite. However, during an optimization that utilizes Hessian updating, and particularly when the initial Hessian is approximate, the sign of some of the eigenvalues of the Hessian may be incorrect. In this case, the pRFO method produces a very ill conditioned shifted Hessian that results in very large steps that are primarily along the eigenvectors with incorrect curvature. Even when the step size is scaled back to a reasonable limit prior to taking the step, it may take the optimizer a number of steps to correct the Hessian sufficiently and continue to make good progress towards the solution. Additionally, it is possible that updates to the Hessian will reduce the overlap of the transition vector with Δb, or increase the overlap of other vectors with Δb, complicating the selection of the transition vector in subsequent steps.

To increase the stability of the transition state optimization, the following divided RFO (dRFO) approach is proposed. First, the eigenvectors of the Hessian are compared against Δb, and any with an overlap of greater than 0.5 are considered to be part of the transition vector space. The augmented Hessian is constructed in each space, and λ_1 is defined as the second most negative eigenvalue of the augmented Hessian in the transition vector space, while λ_2 is defined as the most negative eigenvalue of

the minimization space. In the case where the transition vector space has only one vector, this is identical to the pRFO approach. In addition, the augmented Hessian is constructed and diagonalized in the full space, and the second most negative eigenvalue is used as a shift factor λ_3 that is applied to the entire space.

$$\mathbf{S}_{cpRFO} = -(\lambda_1 + \lambda_3)\sum_{i}^{n_{TS}} v_i v_i^{\mathrm{T}} - (\lambda_2 + \lambda_3)\sum_{i}^{n_{min}} v_i v_i^{\mathrm{T}} \tag{3.13}$$

Just as in the case of the pRFO method, all three of the scale factors in the dRFO method tend towards zero as the magnitude of the gradient is reduced, but the additional correction makes it less likely that the shifted Hessian will be ill-conditioned. This results in an optimization that favors steepest descent in the minimization space over a step along a vector with the wrong curvature when the gradient in the minimization space is still large. A second consequence of the additional correction is that motion in the transition vector space tends to be favored helping the optimizer locate the ridge corresponding to the transition vector before taking large steps downhill, which has been shown[70] to increase the stability of transition state optimizations in the past.

3.3 Implementation and Discussion of Results

The interpolation and optimization methods described in the previous sections will be referred to collectively as the Connectivity Transition State (CTS) method for the following discussion. The CTS method was implemented in the development version of Gaussian 09[71]. Reactant and product structures for the 20 reactions in Figure 1 were optimized on the PM6 and B3LYP[72, 73, 47, 74]/6-31G(d,p) PES. The results of these optimizations were then used as input for the CTS interpolation prior to the transition state search. Table 3.3 summarizes the performance of the CTS method using UFF

during interpolation and optimizing to the transition state on the PM6 PES, while Table 3.4 summarizes the performance of the CTS method using UFF or PM6 during inteprolation and optimizing to the transition state on the B3LYP/6-31G(d,p) PES. In each case, the CTS results are compared with the two input Quadratic Synchronous Transit (QST2) method, and are measured by the total number of optimization steps required for convergence.

The PM6 results in Table 3.3 are shown with analytical Hessians computed at every step (Always), an analytical Hessian computed once at the beginning of the optimization (Once), and estimated force constants without computing an analytical Hessian (Never). In the later two cases, the Hessian approximation is updated at each step using the combined SR1/PSB method proposed by Bofill[28]. The number of steps required for convergence of only the initial minimization phase of the CTS method is also reported separately. For the CTS method, it was determined that computing the Hessian at the beginning of the transition state optimization performed better than computing it at the beginning of the initial minimization phase. Hence, the Once and Never results for CTS both use estimated Hessians for the minimization phase.

The most dramatic difference between the two algorithms is in the interpolation phase. As mentioned previously, the linear interpolation in redundant internal coordinates used by the QST2 method fails to produce a structure in 6 of the 20 reactions using the PM6 minima structures, while the CTS interpolation succeeds for all 20 reactions. An additional 4 QST2 optimizations fail to converge to the correct transition state: DACP2 and Silane follow the incorrect eigenvector and wander away from the region containing the transition state, CN incorrectly converges to a structure with non-zero Cartesian forces due to an error in defining the internal coordinates, and MeOH converges to the transition state for a different reaction. The QST2 method is usually used with reactant-like and product-like structures, which are chosen to more

closely resemble the expected geometry at the transition state than the fully optimized reactants and products. This could account for some of the difficulties the method has with the interpolation and optimization of the reactions in the test set. The CTS method manages to converge to the correct transition state for each reaction at the PM6 level theory when starting from optimized reactant and product structures, while performing similarly to QST2 in most cases where both methods manage to converge.

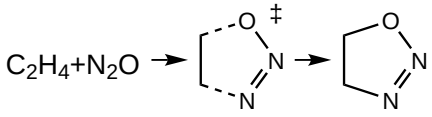
In a number of cases, part or all of the geometry in the CTS guesses using the UFF energy for interpolation had higher symmetry than what is observed in the transition state. For example, the initial CTS structure for 1,3-pentadiene hydrogen transfer has C_{2V} symmetry while the transition state only has C_S symmetry. The initial CTS structure for the Cope rearrangement is a nearly planar ring twisted by ca 35 degrees and has D_2 symmetry, while the transition state has a chair-like geometry with C_{2h} symmetry. Other reactions with symmetries in the initial CTS structures that are not present in the final geometry include the ring closing of 1,3-butadiene, the ene reaction of propene and ethene, and the addition of sulfur dioxide to 1,3-butadiene. The optimizations used to generate the initial CTS structures would seem to favor high symmetry. Since the gradient in the direction of the atomic motion required to break symmetry is zero, it can be difficult for an optimizer to step away from the symmetric geometry and progress towards the transition state geometry, especially when exact Hessian information is not used. This can also be a problem in cases where the initial CTS structure does not necessarily have a higher degree of symmetry than the transition state, but does possess incorrect symmetry that must be broken. This is observed in the CTS structure for the Cope reaction as well as the decomposition of formaldahyde to H_2 to carbon monoxide, which had the initial position of the H_2 rotated by 90 degrees relative to the carbon monoxide, compared to the entirely planar transition structure geometry.

For the majority of the reactions in the test set, the PM6 transition state and the B3LYP/6-31G(d,p) transition state are quite similar, with the exception of the S_N2 reaction which does not have a barrier on the DFT PES. The S_N2 results in Table 3.4 are for the energetically similar reaction of tert-butyl chloride with a fluoride ion. While the transition states were generally similar, the minimum energy structures often differed significantly, particularly in the distance or relative position between bimolecular reactants, with the H_2+CO and Silane reactions showing the largest difference. The relative orientation of the bimolecular reactant structures for both of these reactions at the DFT level provide a much greater challenge for the interpolation than the corresponding minima on the PM6 surface. Additionally some reactions where linear interpolation of the PM6 minima succeeded fail using the DFT minima (C_2NO_2, DACP+eth and ene), while some reactions where linear interpolation of the PM6 minima failed succeed using the DFT minima (DFCP, HF+eth). In most cases, the CTS interpolation using PM6 as the low level energy resulted in a structure that was as good or better for optimization of the DFT transition state than when UFF was used, with DFCP as the only reaction significantly favoring UFF interpolated structure. The PM6 interpolation for the Cope structure does converge more quickly than the UFF interpolation, but it converges to a higher energy boat configuration. The symmetry issues discussed above are not seen during the PM6 interpolation.

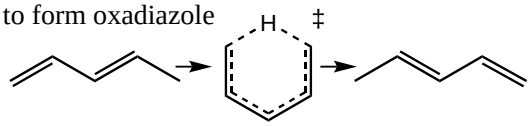
3.4 Summary

The QST2 method implemented in the Gaussian suite of programs is a powerful tool for locating transition state geometries from known reactant and product configurations. The methods described in this paper seek to improve upon the stability and efficiency of locating the transition state by using bonding information to improve the construction of a redundant internal coordinate set to be used for the optimization, and

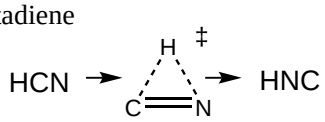
remove the explicit dependence on the reactant and product geometries from the estimation of the initial transition state geometry and the computation of subsequent steps throughout the optimization. Whether exact Hessians are computed or not the CTS method locates transition states with an efficiency comparable to the QST2 method on reactions where both are able to locate the correct transition state, but also succeeds in optimizing most transition states where the QST2 method fails.



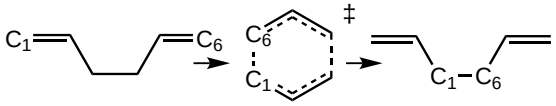
C_2N_2O - addition of nitrous oxide to ethylene to form oxadiazole



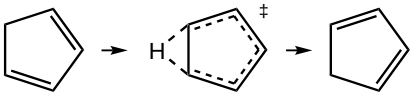
C_5HT - hydrogen transfer from C5 to C1 in 1,3-pentadiene



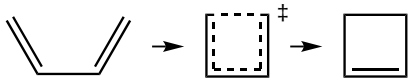
HCN - isomeriszation of hydrogen cyanide



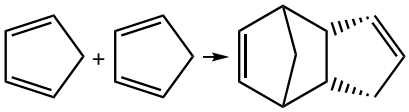
Cope - Cope rearrangment of 1,4-hexadiene



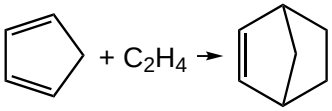
CPHT - hydrogen shift in 1,3-cyclopentadiene



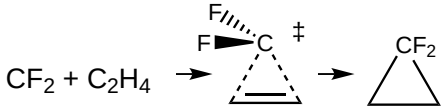
Cyc-But - cyclization of 1,3-butadiene



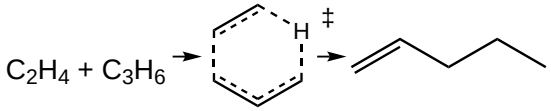
DACP2 - Diels-Alder endo addition of cyclopentadiene to cyclopentadiene



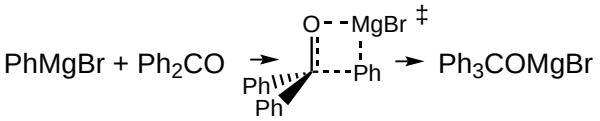
DACP+eth - Diels-Alder addition for cyclopentadiene and ethylene



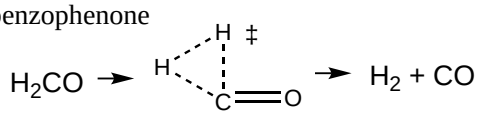
DFCP - difluorocarbene addition to ethylene



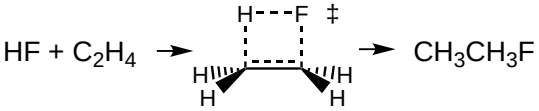
Ene - ene reaction of ethylene and propene



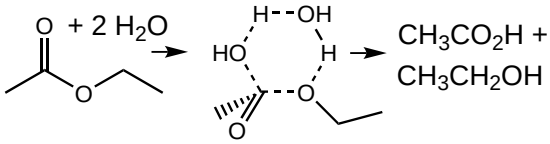
Grignard - addition of phenyl magnesium bromide to benzophenone



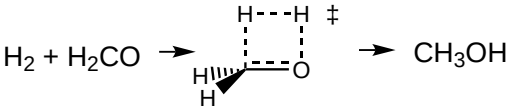
H_2+CO - decomposition of formaldehyde to H_2 + CO



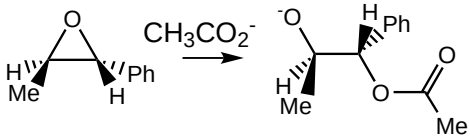
HF+eth - addition of HF to ethylene



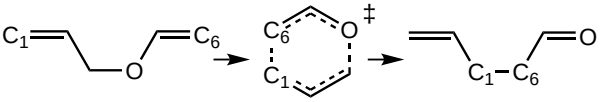
Hydro - water assisted hydrolysis of ethyl acetate



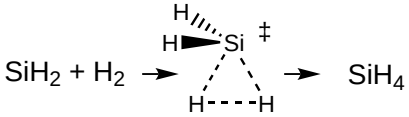
MeOH - addition of H_2 to formaldehyde



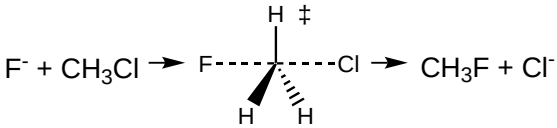
Oxirane - ring opening of 2-methyl-3-phenyloxirane



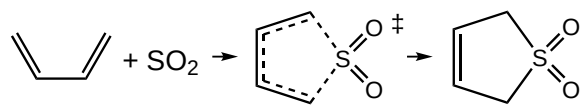
OxyCope - rearrangment to form 1-pentene-5-one



Silane - addition of silylene to H_2



S_N2 - S_N2 reaction of fluoride with methyl chloride



Sulfolene - addition of sulfor dioxide to butadiene

Figure 3.1: Representative test reactions

Table 3.1: Barrier distances versus the bond order at the PM6 transition state geometry

	Barrier			CTS Bond Order[1]			
Reaction	Forward (mH)	Reverse (mH)	Distance[2]	n_{CTS}	mean	max	min
C_2NO_2	4.43	8.68	34%	2	0.32	0.33	0.31
C_5HT	6.44	6.44	50%	2	0.52	0.52	0.52
CN	13.70	11.58	54%	2	0.49	0.66	0.33
Cope	5.99	5.99	50%	2	0.75	0.75	0.75
CPHT	6.13	6.13	50%	2	0.55	0.55	0.55
Cyc-but	7.11	6.36	53%	1	0.45	0.45	0.45
DACP2	5.33	8.88	37%	2	0.43	0.55	0.32
DACP+eth	4.89	8.84	36%	2	0.44	0.44	0.44
DFCP	1.49	7.37	17%	2	0.58	0.77	0.39
Ene	6.16	10.03	38%	3	0.57	0.69	0.42
Grignard	3.74	6.81	35%	2	0.73	1.07	0.39
H_2+CO	12.34	9.32	57%	3	0.59	0.94	0.24
HF+eth	8.23	10.24	45%	3	0.49	0.51	0.45
Hydro	3.23	3.36	49%	6	0.54	0.71	0.40
MeOH	13.39	13.54	50%	3	0.41	0.62	0.23
Oxirane	3.42	2.66	56%	2	0.45	0.46	0.44
Oxycope	5.10	6.80	43%	2	0.62	0.65	0.59
Silane	4.06	8.46	32%	3	0.65	0.92	0.20
S_N2	0.35	6.02	6%	2	0.49	0.73	0.24
Sulfolene	2.65	8.59	24%	2	0.54	0.54	0.54

[1] n_{CTS} is the number of bonds breaking or forming in the reaction

[2] The energetic distance of the barrier is defined as the ratio of the forward barrier height to the sum of the forward and reverse heights. A distance much smaller than 50% can be viewed as an indicator of an early transition state, while a distance much larger than 50% can be viewed as an indicator of a late transition state.

Table 3.2: Comparison of overlap between different tangent approximations and eigenvectors of the Hessian computed at the transition state of the PM6 PES.

	Transition Vector[1]				Other Vector[2]			
Reaction	τ_{lin}	τ_{quad}	τ_{arc}	$\Delta\mathbf{b}$	τ_{lin}	τ_{quad}	τ_{arc}	$\Delta\mathbf{b}$
C_2NO_2	0.69	0.28	0.68	**0.89**	0.38	0.32	0.39	0.38
C_5HT	0.13	0.13	0.13	**0.82**	0.67	0.67	0.67	0.36
CN	0.84	0.77	0.83	**0.91**	0.37	0.37	0.38	0.29
Cope	0.35	0.35	0.35	**0.78**	0.19	0.19	0.19	0.43
CPHT	0.28	0.21	0.28	**0.83**	0.24	0.37	0.24	0.15
Cyc-but	0.13	0.10	0.07	**0.70**	0.59	0.29	0.58	0.52
DACP2	0.20	0.04	0.14	**0.84**	0.56	0.22	0.47	0.36
DACP+eth	0.17	0.03	0.12	**0.83**	0.37	0.16	0.30	0.20
DFCP	0.48	0.52	0.53	**0.89**	0.20	0.22	0.15	0.21
Ene	0.30	0.28	0.35	**0.84**	0.35	0.24	0.33	0.22
Grignard	0.11	0.09	0.11	**0.69**	0.32	0.39	0.32	0.43
H_2+CO	0.60	0.75	0.79	**0.82**	0.25	0.50	0.33	0.40
HF+eth	0.44	0.44	0.44	**0.96**	0.41	0.41	0.41	0.14
Hydro	0.30	0.27	0.30	**0.92**	0.29	0.28	0.29	0.17
MeOH	0.55	0.47	0.54	**0.87**	0.34	0.36	0.30	0.29
Oxirane	0.15	0.18	0.18	**0.84**	0.38	0.26	0.36	0.33
Oxycope	0.26	0.19	0.25	**0.79**	0.26	0.23	0.26	0.34
Silane	0.63	0.56	0.66	**0.72**	0.68	0.48	0.68	0.59
S_N2	**0.96**	0.01	0.83	0.92	0.18	0.74	0.45	0.35
Sulfolene	0.30	0.02	0.23	**0.83**	0.52	0.31	0.37	0.26

[1] Transition vector refers to eigenvector with a corresponding negative eigenvalue. The bold numbers indicate which approximation had the largest overlap with the transition vector.

[2] The other vector columns report the largest overlap with the tangent approximation among the remaining eigenvectors.

Table 3.3: Number of Surface Evaluations Required to Converge at PM6 level of theory

	QST2			CTS (UFF interpolation)				
				Min[1]		Total[1]		
Hessian?	Always	Once	Never	Aways	Never	Always	Once	Never
C_2NO_2	9	13	17	5	5	13	10	20
C_5HT	10	16	15	3	9	11	34	48
CN	fail[2]	fail[2]	fail[2]	4	5	7	8	9
Cope	fail[3]	fail[3]	fail[3]	14	20	19	30	36
CPHT	8	10	16	5	7	7	9	10
Cyc-but	6	10	9	4	5	9	17	21
DACP2	fail[4]	fail[4]	38	5	6	11	12	27
DACP+eth	11	32	24	5	6	7	13	18
DFCP	fail[3]	fail[3]	fail[3]	4	5	7	10	11
Ene	17	35	35	7	9	11	21	30
Grignard	fail[3]	fail[3]	fail[3]	6	10	14	35	78
H_2+CO	10	14	15	7	5	18	23	15
HF+eth	12	19	17	4	5	10	14	17
Hydro	fail[3]	fail[3]	fail[3]	4	5	15	51	42
MeOH	15[5]	20[5]	27[5]	3	5	8	14	14
Oxirane	fail[3]	fail[3]	fail[3]	8	15	19	71	73
Oxycope	fail[3]	fail[3]	fail[3]	8	11	12	17	25
Silane	fail[6]	fail[6]	fail[6]	5	6	11	20	18
S_N2	6	7	9	2	3	7	9	9
Sulfolene	14	30	25	11	7	14	23	24

[1] The Min columns contains only the number of PES evaluations required to converge the minimization phase, while the Total columns contain the PES evaluations required for both the minimization phase and the transition state optimization.

[2] Convergence of the internal coordinate gradient was achieved on a structure with a non-zero Cartesian gradient, indicating an internal coordinate definition error

[3] Linear interpolation failed to produce an initial geometry

[4] Exceeded the default maximum number of optimizaiton steps, which was set as the larger of the number of coordinates being optimized or 100

[5] Optimization converged to the transition state for the decomposition of methanidyloxidanium (CH_2OH_2) to formaldahyde and H_2

[6] Optimizer exit with an error due to very small Hessian eigenvalues

Table 3.4: Number of Surface Evaluations Required to Converge at B3LYP/6-31G(d,p) level of theory

	QST2		CTS			
			PM6 Interpolation		UFF Interpolation	
Hessian?	Once	Never	Once	Never	Once	Never
C_2NO_2	fail[1]	fail[1]	8	9	11	13
C_5HT	11	14	6	7	11	16
CN	fail[2]	fail[2]	6	7	9	10
Cope	fail[1]	fail[1]	7[3]	9[3]	30	37
CPHT	fail[4]	fail[4]	8	9	11	12
Cyc-but	9	9	16	17	17	18
DACP2	45	30	18	31	19	26
DACP+eth	fail[1]	fail[1]	8	10	8	12
DFCP	13	14	31	47	19	19
Ene	fail[1]	fail[1]	9	16	15	16
Grignard	fail[1]	fail[1]	50	51	30	57
H_2+CO	fail[4]	28	fail[4]	18	37	26
HF+eth	fail[1]	fail[1]	9	12	12	15
Hydro	fail[1]	fail[1]	14	24	15	25
MeOH	fail[1]	fail[1]	19	20	12	15
Oxirane	fail[1]	fail[1]	20	41	39	64
Oxycope	fail[1]	fail[1]	11	18	19	25
Silane	24	23	fail[4]	fail[4]	fail[4]	fail[4]
S_N2[5]	25	25	18	30	21	21
Sulfolene	26	27	8	20	32	40

[1] Linear interpolation failed to produce an initial geometry

[2] Convergence of the internal coordinate gradient was achieved on a structure with a non-zero Cartesian gradient, indicating an internal coordinate definition error

[3] Converged to a transition state with a boat configuration, approximately 1.1 mH higher in energy than the chair configuration

[4] Failed to converge within the maximum number of iterations, which was set as the larger of the number of coordinates being optimized or 100

[5] The transition state for methyl chloride + fluoride does not exist at the B3LYP/6-31G(d,p) level of theory. The tert-butyl chloride + fluoride reaction, wihch is energetically similar to the methyl chloride reaction, was used instead.

CHAPTER 4

EXPLORATION OF IMPROVEMENTS TO OPTIMIZATION METHODS

4.1 Introduction

Geometry optimization is an important tool in the computational chemistry toolbox, and has become ubiquitous in modern studies of chemical properties and reactions. There are a wide variety of different algorithms that exist for optimization (see ref [2] for a recent review of methods), with the most common utilizing a combination of quasi-Newton steps in redundant internal coordinates[10, 11], with the Hessian updated by the method of Broyden Fletcher Goldfarb and Shanno (BFGS[22, 23, 24, 25]) when seeking a minimum structure, and either the Powell's symmetric Broyden (PSB[26]) or the symmetric rank 1 (SR1[27]) updates, or some combination of the two[28], is used when attempting to locate a transition state. Approximate, positive definite Hessian matrices[29] are typically used to avoid the relatively large cost of computing the full Hessian exactly. Additionally, sequence acceleration methods such as line searches and direct inversion of the iterative subspace (GDIIS[30]) are also used to reduce the number of potential energy surface (PES) calculations necessary to converge to the desired minimum or transition state structure.

Described below are three new, alternative approaches that may offer additional benefits or improved performance over existing methodologies.

- Flowchart Update - This approach seeks to improve Hessian updating by using different update methods only when they are expected to be well behaved, and falling back to more reliable but less ideal updates when necessary. Additionally, a new modification to the PSB method is used by using scaled displacements to compute the update.

- Scaled RFO method - This approach seeks to improve the use of the rational function optimization method for controlling step size and direction by modifying the shift-matrix to better represent the expected relative stiffness of the bond stretches versus the other coordinates.

- Quasi-rotation method - This is an alternative approach to handling the redundancy in an internal coordinate system. Rather than store the approximate Hessian in the full redundant space, a quasi-rotation matrix is used to rotate the approximate Hessian from the non-redundant space at one point, to the non-redundant space at another. Ideally, this will lead to a more consistent approximation to the Hessian even when the non-redundant space changes over the course of the optimization, and may help improve Hessian updating since the change in the gradient can be expressed entirely in the non-redundant space at one set of coordinates.

A standard method is outlined in section 4.2, and the new methods are detailed in the subsequent sections. The standard method will serve as the benchmark for evaluating the performance of each of the new methods in optimizing a set of 20 molecules with between 10 and 50 heavy atoms (Fig 4.1). This test set was compiled as a more representative collection of molecules being optimized using ab-initio and DFT methods on modern architectures than the commonly used test sets that were developed according to the computer time available in previous decades[65, 12].

4.2 Standard Method

Described below is an optimization algorithm composed of standard methods, which will be used as a control for the purposes of comparing and evaluating new optimization methods. The core of the algorithm is the Newton method applied to the gradient of

the potential energy, and the additional components include the choice of coordinates to optimize, how the Hessian (the matrix of second derivatives of the potential energy surface) is computed or approximated, how the Newton method is modified in order to control step size and direction. While there are a variety of different approaches that have been used in the past, a very common combination of methods include the use of a redundant set of internal coordinates comprised of bond stretches, bends and torsions, approximation of the full Hessian matrix along with the use of updating schemes to improve the approximation over the course of the optimization, and the rational functional optimization and trust region schemes to control the step size and direction.

Geometry extrapolation methods, such as line searches and DIIS, are commonly used to attempt to improve the efficiency of an optimization. In practice, they can also have a significant influence on the stability of an optimization as well. Likewise, Hessian updating methods that consider the change in the gradient at the current geometry relative to many previous optimization steps are employed for similar reasons. In order to better evaluate the merits of the methods described in this chapter on their own terms, no extrapolation methods will be employed, and the Hessian will only be updated using the information at the current and the most recent points.

4.2.1 Coordinate transformation/projection

Coordinate definitions

The standard redundant internal coordinate (RIC) system includes bond stretches between all bonded atoms, angles between pairs of bonds that share an atom, and the dihedral angles between the planes defined by any two angles that share a common bond. Typically, whether or not a pair of atoms is bonded can be determined automatically by comparing the distance between the atoms to a standard bond length,

however, for the present work, the connectivity will be explicitly provided for all examples in order to ensure that the intended coordinate system is used and the behavior of the resulting optimizations does not depend on any arbitrary numerical thresholds for bondedness.

Two additional coordinate definitions that are frequently used in adjunct to the standard stretch, bend and torsion coordinates are out-of-plane bends and linear angle bends. Out-of-plane bends describe how an atom moves relative to the plane defined by 3 other atoms, while linear angle bend coordinates describe the bending motion of 3 atoms which have become nearly or completely linear. Out-of-plane bend coordinates, such as an improper dihedral angle, may provide some benefit to an optimization but are generally unnecessary for structures containing more than 4 atoms as the relevant motions are well described by the redundancy in the standard RIC set. For the present work, no out of plane coordinates are used.

Linear bend coordinates, on the other hand, may be essential when the optimized structures have linear or near-linear angles. Inclusion of standard bend angle coordinates (and the corresponding dihedral angle coordinates) for an angle that has become near linear can be disastrous for an optimization, even when the remaining coordinates in the RIC set are sufficient to fully describe the molecule. Whether or not angles and their corresponding dihedrals can simply be omitted from the RIC definitions depends on how many atoms are bonded to the center atom in the linear angle, and may be possible to do so with as few as 3 bonded atoms. When the central atom has at least 5 bonds, the linear angle coordinates can always be safely omitted since the local structure can be completely defined with bonds and non-linear angles. In the standard method, linear bends will be handled by the approach outlined in section 4.2.2 whenever an angle exceeds 165 degrees and the central atom is bonded to fewer than 5 atoms.

Transformation of gradient/Hessian

Since the derivatives of the PES are computed in terms of the absolute Cartesian coordinates of the atomic centers in a molecule, at each step of the optimization the gradient (and Hessian, if it is computed) must be transformed into internal coordinates. This transformation involves the multiplication of the Cartesian gradient by the generalized inverse of Wilson's B-Matrix, which is defined as the partial derivatives of the RIC with respect to a change in the Cartesian coordinates:

$$B_{ia} = \frac{\partial q_i}{\partial x_a} \tag{4.1}$$

where q_i are the internal coordinates, and x_a are the Cartesian coordinates.

$\mathbf{B}$ is an n_{ric} x n_{crt}, where n_{ric} is the total number of redundant internal coordinates, and n_{crt} is the total number of Cartesian coordinates (3 times the number of atomic centers). The generalized inverse (also known as the Moore-Penrose Pseudoinverse) is a generalization of the concept of an inverse to non-square or singular matrices such that only the non-singular part of the matrix is inverted. Two approaches for computing the pseudo-inverse of the $\mathbf{B}$ matrix, in particular, are through inversion of B-squared ($\mathbf{B}^T\mathbf{B}$), and by singular value decomposition.

For the first approach, it should be recognized that $\mathbf{B}^T\mathbf{B}$ is singular, and so care must be taken when inverting it. So long as the internal coordinates chosen to define $\mathbf{B}$ are sufficient to fully describe the internal motions of the molecule, the singular part of $\mathbf{B}^T\mathbf{B}$ corresponds to the infinitesimal rotations and translations of the molecule. A projector onto this space can be added to $\mathbf{B}^T\mathbf{B}$ without effecting the end result of the pseudo-inverse. This projector, $\mathbf{P}_{TR}$ can be computed as a sum of the projector onto

the individual infinitesimal translation ($\mathbf{t}_i$) and rotation ($\mathbf{r}_i$) vectors

$$\mathbf{P}_{TR} = \sum_{i}^{3} \mathbf{t}_i \mathbf{t}_i^T + \mathbf{r}_i \mathbf{r}_i^T \tag{4.2}$$

where the portion of these vectors corresponding to the kth atom are given by

$$\mathbf{t}_{i,k} = \mathbf{e}_i \tag{4.3}$$

$$\mathbf{r}_{i,k} = \mathbf{x}_k \times \mathbf{e}_i \tag{4.4}$$

where $\times$ denotes the 3-dimesional cross product, $\mathbf{e}_i$ is the unit vector in the ith direction, and $\mathbf{x}_k$ are the 3-dimensional Cartesian coordinates for atom k. These vectors are not orthonormal with respect to each other, but for a non-linear molecule, they fully span and are entirely contained within the 6 instantaneous translation and rotation degrees of freedom for the molecule defined by the coordinates $\mathbf{x}$. The generalized inverse may then be computed as

$$\mathbf{B}^{-} = \left(\mathbf{B}^T\mathbf{B} + \mathbf{P}_{TR}\right)^{-1} \mathbf{B}^T \tag{4.5}$$

The primary benefit of computing the generalized inverse this way is that $\left(\mathbf{B}^T\mathbf{B} + \mathbf{P}_{TR}\right)^{-1}$ allows the use of iterative inversion methods for when the size of the molecule is too large to invert the matrix directly and/or store the result. For the present work, however, the generalized inverse will be computed using singular value decomposition (SVD). The SVD of $\mathbf{B}$ is given by $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}$ is a square, unitary matrix of size n_{ric}, $\mathbf{V}$ is a unitary matrix of size n_{crt}, and $\mathbf{\Sigma}$ is a matrix of the appropriate dimensions that is zero everywhere except for the first $n_{crt} - 6$ diagonal

elements. The generalized inverse may be computed as

$$\mathbf{B}^{-} = \mathbf{V}\boldsymbol{\Sigma}^{-}\mathbf{U}^{T} \tag{4.6}$$

where $\boldsymbol{\Sigma}^{-}$ is the transpose of $\boldsymbol{\Sigma}$, with the non-zero diagonal elements replaced by their inverse (i.e. $\left[\boldsymbol{\Sigma}^{-}\right]_{ii} = {}^{1}/_{\Sigma_{ii}}$ if $\Sigma_{ii} \neq 0$). Whether eq 4.5 or eq 4.6 are used, the same matrix should result. The primary benefit of using eq 4.6 is that the locally non-redundant *active space* for the current geometry (the $n_{act} = n_{crt} - 6$ orthogonal linear combinations of the redundant internal coordinates in which $\mathbf{B}$ is defined for a given geometry) is computed as a consequence of doing the SVD. The active space is defined as the first n_{act} rows of $\mathbf{U}$.

With $\mathbf{B}^{-}$ defined, the typical approach for transformation of the gradient and Hessian begins with an algebraic rearrangement of the chain rule definition of the gradient (using x and q subscripts to differentiate between Cartesian and redundant internal coordinate gradients, respectively)

$$\mathbf{g}_x = \mathbf{g}_q^T\mathbf{B} \rightarrow \mathbf{g}_q = \mathbf{g}_x^T\left(\mathbf{B}^{-}\right) \tag{4.7}$$

A similar approach is used to transform the Hessian

$$\mathbf{H}_q = \left(\mathbf{B}^{-}\right)\left(\mathbf{H}_x - \mathbf{g}_q^T\frac{\partial\mathbf{B}}{\partial\mathbf{x}}\right)\left(\mathbf{B}^{-}\right)^T \tag{4.8}$$

Projection into non-redundant space

When using a redundant coordinate system, the total number of coordinates being optimized exceeds the number of internal degrees of freedom, and this must be accounted for when computing a step. Constraints or a penalty function approach could

be used to compute a valid step in the full set of redundant coordinates, but here a projection method will be used. The active space described in the previous section may be used to project the gradient $\mathbf{g}_r = \mathbf{U}_{act}\mathbf{g}_q$ and Hessian $\mathbf{H}_r = \mathbf{U}_{act}\mathbf{H}_q\mathbf{U}_{act}^T$ into a reduced space. The step, $\mathbf{\Delta q}_r$, is computed in this space, and then transformed back to the full $\mathbf{\Delta q}_0 = \mathbf{U}_{act}^T\mathbf{\Delta q}_r$

Back-transformation of step

Due to the curvilinear relationship between the redundant internal and Cartesian coordinates, the back-transformation has to be done iteratively. The conceptual approach used in the present work is to minimize the difference between the current geometry $\mathbf{q}(\mathbf{x})$ and a goal geometry, in redundant internal coordinates $\mathbf{q}_g = \mathbf{q}_0 + \mathbf{\Delta q}_0$, where $\mathbf{q}_0$ is the internal coordinate geometry at the current iteration, and $\mathbf{\Delta q}_0$ is the computed step. This can be thought of as the minimization of the following functional

$$F(\mathbf{x} + \mathbf{\Delta x}) = \frac{1}{2}(\mathbf{\Delta q})^2 = \frac{1}{2}(\mathbf{q}_g - \mathbf{q}(\mathbf{x} + \mathbf{\Delta x}))^2 \tag{4.9}$$

with respect to $\mathbf{\Delta x}$, starting with $\mathbf{x} = \mathbf{x}_0$ where $\mathbf{x}_0$ is Cartesian geometry at the current step. This gives the typical iterate for $\mathbf{\Delta x}$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{B}_i^- \mathbf{\Delta q}_i \tag{4.10}$$

This process continues until the RMS of $\mathbf{\Delta x}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$ is less than 10^{-6} bohr.

4.2.2 Linear Bends

Definition of terms

For notational clarity, a few items will be defined before describing how the linear bend coordinates will be constructed and transformed. Subscripts will be used to

differentiate between atoms, x_i, $r_{ij} > 0$, $0 \leq \theta_{ijk} \leq \pi$ and $-\pi \leq \phi_{ijkl} \leq \pi$ will be used to indicate the Cartesian coordinates, bond lengths, bend angle, and torsion angle. The atoms involved in the linear angle will be indicated by numbers 1 through 3 (i.e. the angle which is nearly/completely linear is θ_{123}), the dummy atom will use a subscript d, and n and m will be used to indicate any atoms bonded to 1 and 3, respectively.

Placement of the dummy atom

If θ_{123} is not completely linear, then the plane in which the angle bends may be defined by r_{12} and r_{23}, while if it is linear, the choice of bend direction is arbitrary. For consistency and convenience, the bend direction may be chosen such that either ϕ_{n12d} or ϕ_{d23m} is 0 for some n or m. Once the plane is defined, the dummy atom is placed such that $\theta_{12d} = \theta_{32d}$ and r_{2d} is equal to a constant value. In the present implementation, this is set to 2 bohr since that is a fairly typical bond length, but the choice is more or less arbitrary.

Definition of coordinate set to handle linear bends

With the dummy atom placed, the optimization is carried out by modifying the coordinate definitions and constraining three coordinates per linear bend to define the location of the dummy atom relative to the rest of the molecule. The standard approach is used to construct the coordinate set beginning with all of the stretches corresponding to the bonded atoms, and angle bends between pairs of stretches that share a common atom. The angle θ_{123} is redefined as the equivalent sum $\theta_{12d}+\theta_{32d}$. The dihedral angles are included for all pairs of angle bend coordinates (not including θ_{123}) that share a bond, including all of the dihedrals ϕ_{n12d} and ϕ_{d23m}, and the improper dihedral ϕ_{12d3}.

The constrained coordinates will always include r_{2d}, $\theta_{12d} - \theta_{32d}$, and a dihedral, but the choice of dihedral depends on the structure. For the vast majority of cases,

the choice of any ϕ_{n12d} or ϕ_{d23m} is appropriate, with special care taken when there are multiple adjacent linear bends (i.e. 4 or more bonded, colinear atoms) to ensure that one dihedral for each dummy atom is included. If the molecule contains only 3 atoms, the improper dihedral ϕ_{12d3} can be constrained as the bend direction is completely arbitrary for a potential that only depends on the internal coordinates.

Coordinate transformation/back-transformation

Let $\mathbf{B}_{opt}$ be the B-matrix for the redundant internal coordinate set defined in the previous section, and $\mathbf{B}_{cns}$ be the B-matrix for the set of coordinates that will be used to constrain the motion of the dummy atom. The transformation is carried out using an otherwise identical approach to the standard unconstrained transformation, using a projected B-matrix $\mathbf{B}_{prj} = \mathbf{B}_{opt} - \mathbf{B}_{opt}\mathbf{B}_{cns}^{-}\mathbf{B}_{cns}$, and assuming that the gradient and Hessian elements corresponding to the dummy atoms are all 0. $\mathbf{B}_{prj}$ has the correct number of non-singular values ($3*n_{atoms}$) for the original system. If analytical Hessians are used, the B-matrix derivatives are similarly projected: for internal coordinate i, $\partial\left[\mathbf{B}_{prj}\right]_i = \left(\mathbf{I} - \mathbf{B}_{cns}^{-}\mathbf{B}_{cns}\right)\partial\left[\mathbf{B}_{opt}\right]_i\left(\mathbf{I} - \mathbf{B}_{cns}^{-}\mathbf{B}_{cns}\right)$.

Once a step is computed, the back-transformation is done in an unconstrained fashion, with the goal values for the constrained coordinates set according to their current values. A second iterative back-transformation is used, if necessary, to reimpose the dihedral constraint since there may be some redundancy between the dihedral constraint chosen and dihedral coordinates used in the optimization. During the second iterative back-transformation, only the dummy atom is allowed to move.

4.2.3 Step computation/constraint

RFO

The rational function optimization[54] (RFO) method begins by approximating the PES as a rational function:

$$E_{RF}\left(\mathbf{q}_0 + \mathbf{\Delta q}\right) = E_0 + \frac{\mathbf{\Delta q}^T \mathbf{g}_0 + \frac{1}{2}\mathbf{\Delta q}^T \mathbf{H}_0 \mathbf{\Delta q}}{1 + \mathbf{\Delta q}^T \mathbf{\Delta q}} \tag{4.11}$$

Eq 4.11 has the property that, for a non-zero gradient, there exactly $n+1$ stationary points on the surface, where n is the number of coordinates defining the surface, with each stationary point having a different curvature (determined by the number of negative eigenvalues in the second derivative of E_{RF}). By stepping towards the stationary point on the rational function surface with the desired curvature, an optimization may be directed to a particular solution even when the number of negative eigenvalues in $\mathbf{H}_0$ is incorrect. The stationary condition (the value of $\mathbf{\Delta q}$ where $\partial E_{RF}/\partial \mathbf{\Delta q}$ is zero) for all $n+1$ stationary points is given by diagonalizing the augmented Hessian

$$\mathbf{H}_{aug} = \begin{bmatrix} \mathbf{H} & \mathbf{g} \\ \mathbf{g}^T & 0 \end{bmatrix} \tag{4.12}$$

The step may be computed using either the mth eigenvector (v_m) or eigenvalue (λ_m) of the augmented Hessian.

$$\mathbf{\Delta q}_{RFO,m} = \frac{\mathbf{v}_{m,1..n}}{v_{m,n+1}} \tag{4.13}$$

$$\mathbf{\Delta q}_{RFO,m} = -\left(\mathbf{H} - \lambda_m \mathbf{I}\right)^{-1} \mathbf{g} \tag{4.14}$$

Both approaches result in the same displacement, and the resulting step satisfies the following relationship

$$\lambda_m = \left(\mathbf{\Delta q}_{RFO,m}\right)^T \mathbf{g} \tag{4.15}$$

For minimizations, the most negative eigenvalue of the augmented Hessian is used. For transition states, even though the second most negative eigenvalue will result in a shifted Hessian with a single negative eigenvalue, the partitioned RFO (pRFO[54]) approach tends to result in a more stable optimization behavior. The pRFO approach involves searching for a minimum in $n-1$ directions (usually selected to be eigenvectors of the Hessian), and a maximum in the remaining direction. This is done by combining an RFO minimization step in the $n-1$ space, and an RFO maximization step in the remaining space.

The RFO method provides a means to compute a step in the correct direction for an optimization even when the Hessian has the incorrect number of negative eigenvalues (0 for minima, 1 for transition states). Often times, however, when the geometry is far from the solution, and especially when the Hessian data is approximate/updated, the computed step size may be unreasonably large and will need to be scaled back to avoid over-shooting the solution. For minimizations, an adaptive step size approach is frequently used, and will be included in the standard method. The adaptive step size, or trust region, is initialized to a modest value (0.5 bohr or rad for the present work), and at each step of the optimization, the predicted quadratic change in energy ($\mathbf{\Delta} E_{quad} = \mathbf{\Delta q}^T\mathbf{g} + 0.5\mathbf{\Delta q}^T\mathbf{H}\mathbf{\Delta q}$) is compared to the actual change in the energy. When the ratio of $\mathbf{\Delta} E_{quad}$ and the actual change in energy is greater than 75% and the length of the step is at least 80% of the current trust region, the trust region is doubled. If the ratio of $\mathbf{\Delta} E_{quad}$ and the actual change in energy is less than 25%, the trust region is reduced to a quarter of the length of the previous step.

4.2.4 Hessian Update

To facilitate the discussion of different Hessian update formulas, the following standard notations for the displacement $\mathbf{s} = \mathbf{\Delta q}$, change in gradient $\mathbf{y} = \mathbf{\Delta g}_q$, and quadratic error $\mathbf{z} = \mathbf{y} - \mathbf{Hs}$ will be used.

$$\mathbf{\Delta H}_{BFGS} = \frac{\mathbf{y}\mathbf{y}^T}{\mathbf{y}^T\mathbf{s}} - \frac{\mathbf{sHHs}^T}{\mathbf{sHs}} \tag{4.16}$$

$$\mathbf{\Delta H}_{SR1} = \frac{\mathbf{z}\mathbf{z}^T}{\mathbf{z}^T\mathbf{s}} \tag{4.17}$$

$$\mathbf{\Delta H}_{PSB} = \frac{\mathbf{s}\mathbf{z}^T + \mathbf{z}\mathbf{s}^T}{\mathbf{s}^T\mathbf{s}} - \mathbf{s}^T\mathbf{z}\frac{\mathbf{s}\mathbf{s}^T}{\left(\mathbf{s}^T\mathbf{s}\right)^2} \tag{4.18}$$

$$\mathbf{\Delta H}_{MSP} = \phi_{MSP}\mathbf{\Delta H}_{SR1} + \left(1 - \phi_{MSP}\right)\mathbf{\Delta H}_{PSB} \tag{4.19}$$

$$\phi_{MSP} = \frac{\left(\mathbf{s}^T\mathbf{z}\right)^2}{\left(\mathbf{z}^T\mathbf{z}\right)\left(\mathbf{s}^T\mathbf{s}\right)} \tag{4.20}$$

When attempting to locate a minimum energy structure, the BFGS update (eq 4.16) is widely considered to be the gold standard as it is generally quite stable, so long as $\mathbf{s}^T\mathbf{y}$ is positive and the current approximation to the Hessian is positive definite. For transition state optimizations, which require the Hessian to be indefinite with one negative eigenvalue, there is not such a clear cut winner. The SR1 method (eq 4.17, also referred to as the MS or Murtagh-Sargeant update) tends to produce the most reasonable updates, particularly when the change in the Hessian is large (for example, when the approximation is poor or the change in the gradient is large), but it can suffer from stability issues when $\mathbf{s}^T\mathbf{z}$ is small relative to $\mathbf{z}^T\mathbf{z}$. The PSB method (eq 4.18),

on the other hand, is very stable, but the updates tend to be significantly poorer as they increase in magnitude. A combination of these two updates, the MSP method (eq 4.19), was proposed by Bofill[28]. This combined update attempts to leverage the pros and cons of the SR1 and PSB methods by blending them together according to the overlap of the quadratic error and the displacement (eq 4.20).

4.2.5 Convergence

An optimization is considered converged when the RMS change in $\mathbf{\Delta q}$ is less than 1.2×10^{-3}, the maximum absolute element in $\mathbf{\Delta q}$ is less than 1.8×10^{-3}, the RMS of $\mathbf{g}_r$ is less than 1.5×10^{-4}, and the maximum absolute element in $\mathbf{g}_r$ is less than 4.5×10^{-4}.

4.3 Flowchart Update

4.3.1 Motivation

It's fairly well understood that the BFGS (eqn 4.16) and SR1 (eqn 4.17) updates can become numerically unstable when $\mathbf{y}^T\mathbf{s}$ and $\mathbf{z}^T\mathbf{s}$ are small, respectively. The PSB update (eqn 4.18), on the other hand, is always numerically stable since the only quantity in the denominator ($\mathbf{s}^T\mathbf{s}$) is always non-zero for a finite step, but may have undesirable properties when the quadratic error is large. After observing that, in general, the SR1 update produced the most reasonable results when $\mathbf{z}^T\mathbf{s}$ is less than zero, the following flowchart method was developed.

4.3.2 Method

Flowchart Method

1. If $\frac{\mathbf{z}^T\mathbf{s}}{|\mathbf{z}||\mathbf{s}|} < -0.1$, use the SR1 update

2. If $\frac{\mathbf{y}^T\mathbf{s}}{|\mathbf{y}||\mathbf{s}|} > 0.1$, use the BFGS update

3. Otherwise, use PSB method

This has the benefit of attempting to use the SR1 and BFGS methods, which are often far superior to the PSB method for minimization, as much as possible, but relying on the numerical stability of PSB when necessary.

SSB Method

All Hessian update methods are based upon computing a correction to the Hessian which satisfies the secant equation

$$(\mathbf{H} + \mathbf{\Delta H})\,\mathbf{s} = \mathbf{y} \tag{4.21}$$

Whenever the dimensionality of the PES is greater than 1, this equation is under determined for the correction ($\mathbf{\Delta H}$) and the imposition of different constraints, such as the requirement that the correction be symmetric, and that it has a minimum size according to some metric, leads to the different update formula. Since all update methods satisfy eq 4.21, the difference in performance between the methods must depend on how they treat the remaining space. For SR1 and BFGS, the vectors used to construct the update tend to resemble gradient change terms ($\mathbf{y}$, $\mathbf{Hs}$ or $\mathbf{z}$), while PSB is constructed primarily with displacements. Similarly, the denominators in the SR1/BFGS updates are scalar products of the displacement and a gradient change terms, while the denominators in the PSB update are scalar products of the displacement alone. If these features play a role in the improved behavior observed with the SR1/BFGS updates, then perhaps the PSB update can be modified to produce more reasonable updates as well.

A general "symmetric rank 2" update can be constructed using any vector $\mathbf{v}$ as follows

$$\mathbf{\Delta H}_{SR2} = \frac{\mathbf{v}\mathbf{z}^T + \mathbf{z}\mathbf{v}^T}{\mathbf{v}^T\mathbf{s}} - \mathbf{s}^T\mathbf{z}\frac{\mathbf{v}\mathbf{v}^T}{\left(\mathbf{v}^T\mathbf{s}\right)^2} \tag{4.22}$$

This update is valid for any choice of $\mathbf{v}$ that has a non-zero overlap with $\mathbf{s}$, with $\mathbf{v} = \mathbf{s}$ giving the PSB update, and $\mathbf{v} = \mathbf{z}$ giving the SR1 update. One simple way to choose a sensible $\mathbf{v}$ that is both gradient change like and based upon the current geometry that will never have a near-zero overlap with the displacement $\mathbf{s}$ is to simply scale $\mathbf{s}$ by a positive definite matrix $\mathbf{v} = \mathbf{M}\mathbf{s}$. The most convenient choice of $\mathbf{M}$ would be to use the approximate Hessian that's used to initialize a quasi-Newton optimization. The use of this "scaled" symmetric Broyden (SSB) update in the flowchart method will also be examined.

4.4 Scaled RFO method

4.4.1 Motivation

The shifted (quasi-)Newton methods control the direction and magnitude of a (quasi-)Newton step by a shift of the Hessian eigenvalues. When the shift factor is large enough to dominate the Hessian, the result is a scaled steepest descent (or shallowest ascent, depending on the sign of the shift factor). This tends to be a poor step on a chemical PES, since the vibrational modes in a molecule tend to be a combination of both soft angular/torsional and stiff bond stretch coordinates, which leads to oscillations and poor convergence behavior. Standard optimization methods using a shifted Newton method (such as RFO) produce a shift factor that is small enough that the Hessian is dominant, and the vibrational modes are scaled appropriately to avoid the oscillatory behavior. However, it is still worth investigating whether a superior

method of controlling the direction and magnitude of the (quasi-)Newton step might exist, and the scaled RFO method is the result of that investigation.

4.4.2 Method

The scaled RFO method is functionally similar to the standard RFO method, with the difference being that the displacement term in the denominator of the energy function 4.11 is scaled

$$E_{SRF}\left(\mathbf{q}_0 + \mathbf{\Delta q}\right) = E_0 + \frac{\mathbf{\Delta q}^T\mathbf{g}_0 + \frac{1}{2}\mathbf{\Delta q}^T\mathbf{H}_0\mathbf{\Delta q}}{1 + \mathbf{\Delta q}^T\mathbf{S}\mathbf{\Delta q}} \tag{4.23}$$

The matrix $\mathbf{S}$ should be positive definite, and should better account for the difference in stiffness between the coordinates. Additionally, to reproduce a similar optimization behavior to what is observed in the RFO method, $\mathbf{S}$ should be normalized so that $\mathbf{\Delta q}^T\mathbf{S}\mathbf{\Delta q} = \mathbf{\Delta q}^T\mathbf{\Delta q}$. This ensures that the only function of the $\mathbf{S}$ matrix is to change the contribution from each coordinate in the shifted matrix relative to one another without increasing or decreasing the magnitude of the shift itself. These requirements are met by using an approximate, positive definite Hessian $\tilde{\mathbf{H}}$ such as those used to initialize the Hessian in a quasi-Newton optimization, with the normalization factor computed as the inverse determinant.

$$\mathbf{S} = \frac{\widetilde{\mathbf{H}}_r}{\left|\widetilde{\mathbf{H}}_r\right|^{1/n_{act}}} \tag{4.24}$$

Löwdin orthogonalization[75] can be used to scale the Hessian, gradient and computed step, allowing for the computation of the shift parameter in the standard way.

$$\mathbf{\Delta q}_{SRFO} = \mathbf{S}^{1/2}\left(\mathbf{S}^{1/2}\mathbf{H}_r\mathbf{S}^{1/2} - \lambda_{RFO}\mathbf{I}\right)^{-1}\mathbf{S}^{1/2}\mathbf{g}_r \tag{4.25}$$

4.5 Quasi-rotation internal coordinate propagation

4.5.1 Motivation

Regardless of the coordinate system chosen, the number of internal degrees of freedom defining a molecular geometry is fixed at $n_{act} = n_{crt} - 6$ for non-linear molecules. Whenever more than n_{act} primitive coordinates are used to define the molecular geometry, a local basis of n_{act} linear combinations of the primitive coordinates completely define the allowed internal motions of the molecule, and the remaining coordinates orthogonal to that basis are redundant. This basis is considered local because it depends on the corresponding geometry, in the case of Cartesian coordinates the rotational orientation defines the direction of the overall infinitesimal rotations, while with redundant internal coordinates, the local basis is given by the left singular vectors of the B-matrix, which is a function of the Cartesian coordinates.

Presented below is a proposed method for transforming the derivatives in the local basis for one geometry into the local basis for another geometry. The immediate benefits for doing this are that it reduces the amount of space required to store approximate Hessian information down to the amount required for a symmetric n_{act} matrix, and that it may help keep approximate Hessian information relevant over longer optimizations when the local space changes dramatically. The transformation is constructed by projecting the B-matrix of a previous geometry into the local basis at the current geometry, and the resulting n_{act} x n_{act} matrix will have an unsigned determinant of approximately, but not less than, 1. For this reason, the transformation is referred to as a quasi-rotation coordinate propagation method.

4.5.2 Method

Instead of updating the Hessian in the full redundant space and then projecting into the non-redundant space at every step of the optimization, the Hessian may be stored and updated in only the active space. In order to account for how the active space changes from point to point, a quasi-rotation scheme is used. If $\mathbf{B}^{old}$ and $\mathbf{B}^{new}$ are the B-matrix computed at the previous and current point, respectively, while $\mathbf{U}_{act}^{old}$ and $\mathbf{U}_{act}^{new}$ are the active space at the previous and current point, the quasi-rotation matrix that transforms a matrix or vector from the active space at the previous point to the active space at the current point is given by

$$\mathbf{M}^{rot} = \mathbf{U}_{act}^{old}\mathbf{B}^{old}\left(\mathbf{U}_{act}^{new}\mathbf{B}^{old}\right)^{-1} \tag{4.26}$$

The rotated Hessian at the new point may be computed as follows

$$\mathbf{H}_{act}^{rot} = \left(\mathbf{M}^{rot}\right)^{T}\mathbf{H}_{act}^{old}\mathbf{M}^{rot} \tag{4.27}$$

This may then be updated by using modified $\mathbf{\Delta q}$ and $\mathbf{\Delta g}$ vectors

$$\mathbf{\Delta q}^{prj} = \mathbf{U}_{act}^{new}\left(\mathbf{q}^{new} - \mathbf{q}^{old}\right) \tag{4.28}$$

$$\mathbf{\Delta g}^{rot} = \mathbf{g}_{r}^{new} - \left(\mathbf{M}^{rot}\right)^{T}\mathbf{g}_{r}^{old} \tag{4.29}$$

The projected displacement will not be 100% accurate, as some of the displacement may exist in the redundant space at the new point. However, this amount is typically small relative to the magnitude of the displacement vector, and the BFGS and SR1 updates only modify space that is non-zero in $\mathbf{\Delta g}$ and $\mathbf{H\Delta q}$.

4.6 Results and Discussion

To compare the effect each method has on an optimization relative to the standard methods, the following algorithms were constructed from the standard approach in section 4.2:

Standard The reference approach

FlowPSB Replace BFGS updating with the flowchart update, using the PSB update as the fallback option

FlowSSB Replace BFGS updating with the flowchart update, using the SSB update as the fallback option

SRFO Compute the step using the scaled RFO approach

Rotcrd Store only the non-redundant n_{act} x n_{act} Hessian, using the quasi-rotation matrix to propagate the Hessian and gradient at the previous point prior to applying the update

These methods were implemented in a Mathematica[76] program, using Gaussian 09[71] to evaluate the potential energy surface. Table 4.1 shows the number of iterations, consisting of an energy and gradient calculation, that each method required to achieve the convergence criteria listed in section 4.2.5 for the examples in the test set. HF/3-21G was used to compute the energies in every case but the magnesium porphin and vitamin C optimizations, which used B3LYP[72, 73, 47, 74] with a 6-31G(d,p) basis set. The total number of iterations required to converge all of the structures in the test set is listed for each method. The rotcrd method was the only one that required more iterations (701) than the standard method (689) to optimize the test set. In general, the unsigned determinant of the quasi-rotation matrix was less than 1.05 even

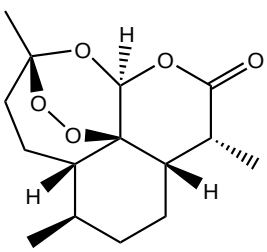
for large steps, however, when the BFGS update produced an indefinite Hessian, the subsequent step not only tended to increase the energy, but the resulting quasi-rotation matrix had a determinant much larger than previous or subsequent steps, in some cases being as large as 1.2. Since corresponding poor updates were observed in the standard algorithm, it suggests that changes to the local basis may play some role in these poor updates, and that a large determinant of the quasi-rotation matrix may have some use in rejecting what's likely to be a poor step prior to evaluating the energy/gradient. Additionally, the results here indicate that the rotcrd method does not appear to introduce too significant of an error and new strategies that utilize the rotcrd approach, such as methods involving a change to the coordinate set mid-optimization, should be investigated.

Both of the flowchart methods and the SRFO method required fewer iterations to optimize all of the structures in the test set than the standard method. Only 8 of the 20 optimizations needed to fall back to the symmetric Broyden update in the flowchart method, so the differences in performance between the FlowPSB and FlowSSB methods are limited to the differences in those 8 structures with neither approach doing consistently better than the other. Even when the symmetric Broyden update was needed, it was only used once or twice during the course of the optimization, and so the BFGS and SR1 updates were well suited to these problems. It's possible that greater differences between the BFGS and the flowchart updates would be observed when the surface is less regular than ab-initio or DFT surfaces (for example, using semi-empirical energies, or solvation models), and future investigation is warranted.

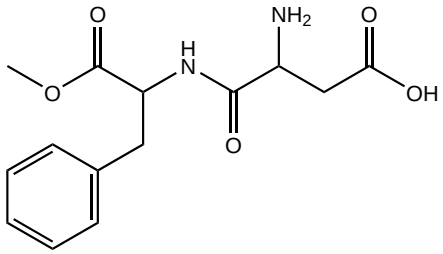
The SRFO method had the best performance of the 5, requiring 45 fewer iterations to converge the test set than the standard approach. The SRFO method was also the only method that converged to different structures in six of the cases. When the initial structure is high in energy and far from a minimum, there may be multiple minimum

energy wells that may be accessible, each containing a different stable conformer of the reaction. Since the SRFO effectively emphasizes changes to dihedral and valence angles early in the optimization, it is reasonable to expect that it will occasionally wind up in a different well than the standard RFO approach, even when using the same initial Hessian and same Hessian update formula. The molecules where the SRFO method had the largest improvement over the standard method (Azadirachtin, Bisphenol A, EASC and Inosine), the SRFO and Standard methods converged to the same structure. Likewise, the only case where the SRFO method needed a significant number of additional iterations to converge (Ochratoxin A), the SRFO method produced a different structure.

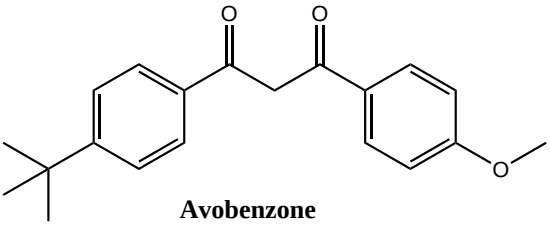
Figure 4.1: Structures for test set



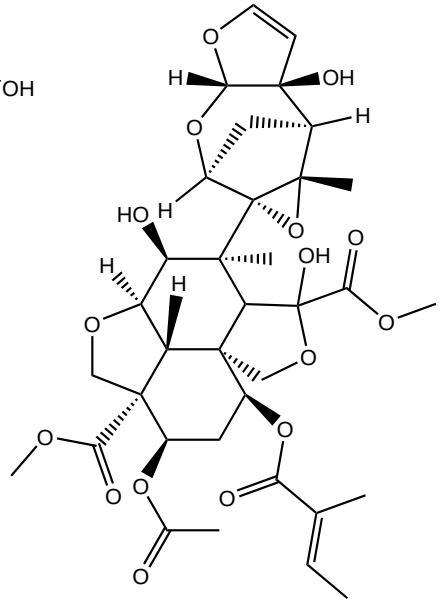
Artemisinin



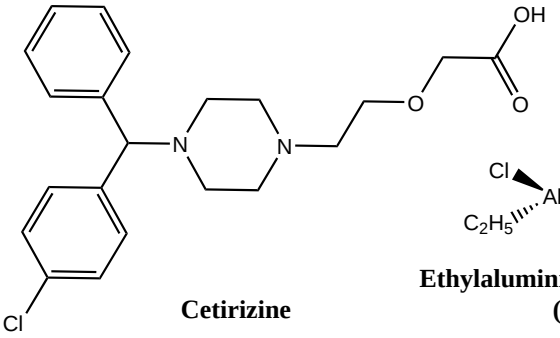
Aspartame



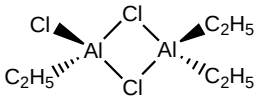
Avobenzone



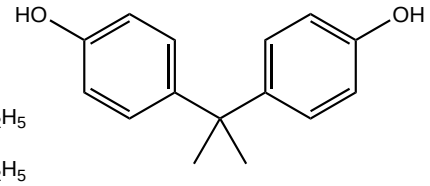
Azadirachtin



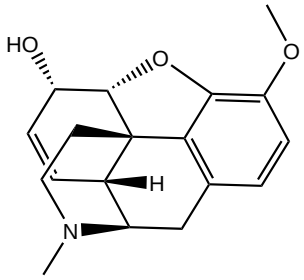
Cetirizine



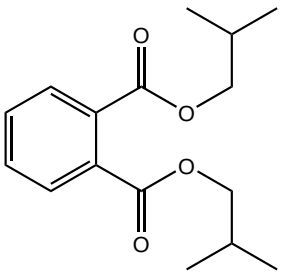
Ethylaluminium Sesquichloride (EASC)



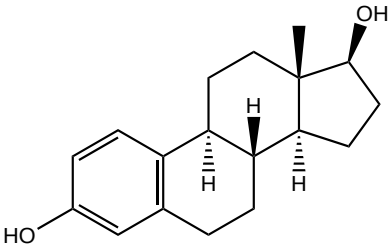
Bisphenol A



Codeine

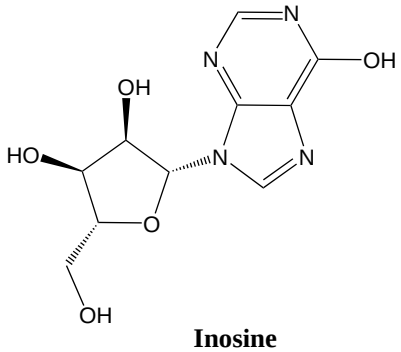


Diisobutylphthalate

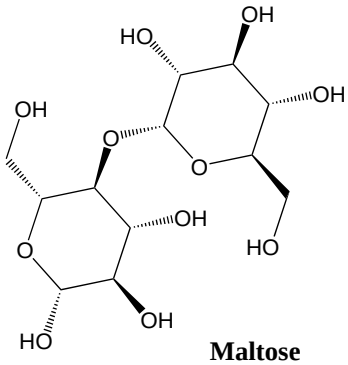


Estradiol

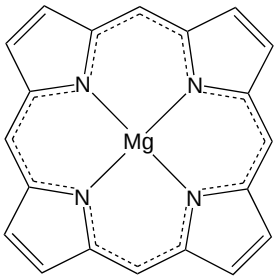
Figure 4.1: Structures for test set (cont)



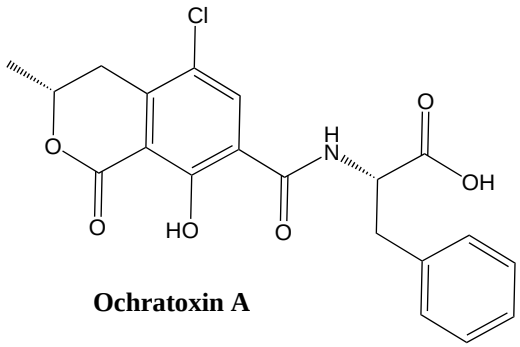
Inosine



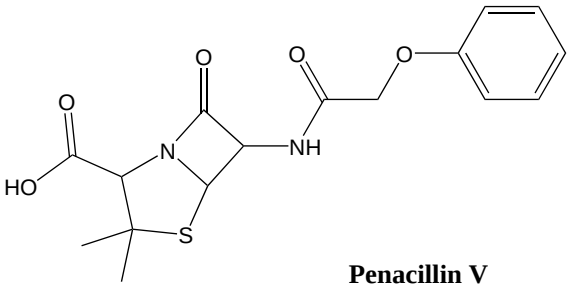
Maltose



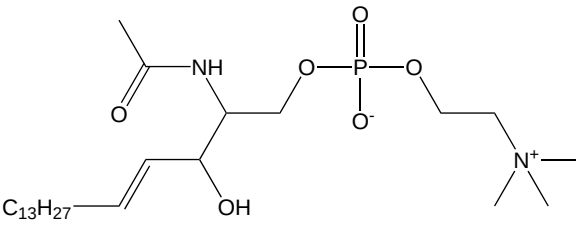
Magnesium Porphin



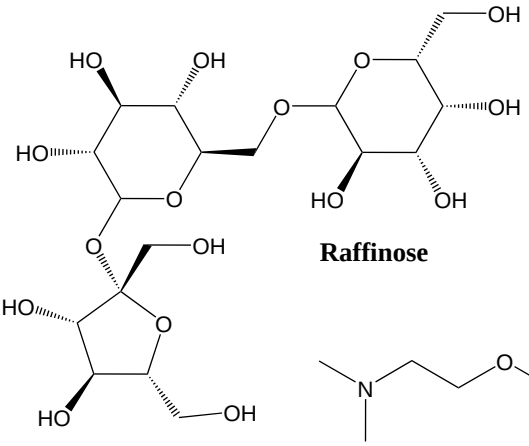
Ochratoxin A



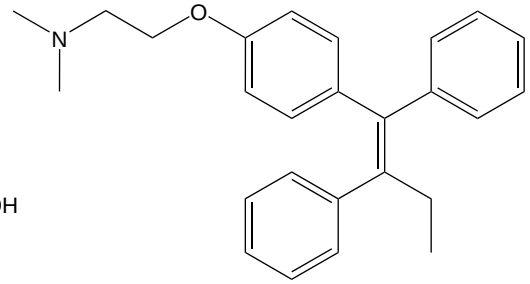
Penacillin V



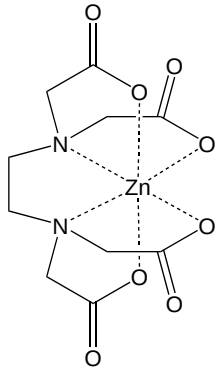
Sphingomyelin



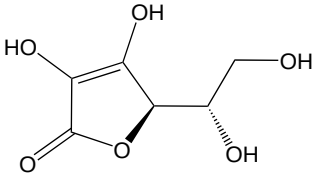
Raffinose



Tamoxifen



Zinc EDTA



Vitamin C

Table 4.1: Number of iterations required to converge to a minimum structure

	Standard	FlowPSB	FlowSSB	SRFO	Rotcrd
Artemisinin	24	21	21	24	25
Aspartame	30	30	30[1]	28[2]	27
Azadirachtin	77	73	74	65	72
Avobenzone	43	41	42	43	47
Bisphenol A	31	32	31	21	30
Cetirizine	35	37	40	30	39
Codeine	18	18	18[1]	18[3]	19
Diisobutylphthalate	27	27	27[1]	28[2]	29
Estradiol	21	21	21[1]	19	21
EASC	40	27	25	24	31
Inosine	41	46	42	34	41
Maltose	35	36	37	36	37
Mg porphin	14	24	24[1]	16	14
Ochratoxin A	31	31	31[1]	43[3]	31
Penicillin V	31	32	32[1]	30[2]	32
Raffinose	53	55	55[1]	54[3]	55
Sphingomyelin	56	48	48[1]	51	67
Tamoxifen	35	35	35[1]	35	36
Vitamin C	18	18	18[1]	18	17
Zn EDTA	29	29	29[1]	27	31
Total	689	681	680	644	701

[1] The symmetric Broyden update was never used in the FlowPSB optimization, so the optimization behavior of the FlowSSB method is identical

[2] The SRFO optimization converged to a higher energy structure than the other methods

[3] The SRFO optimization converged to a lower energy structure than the other methods

CHAPTER 5

PATH OPTIMIZATION BY A VARIATIONAL REACTION COORDINATE METHOD

5.1 Introduction

With recent developments in reaction path following[19, 20, 31], the steepest descent reaction path (SDRP) can be readily determined by walking downhill from a transition state on the potential energy. At any point along this pathway, the energy orthogonal to the path is a local minimum, and so it provides a first-order approximation of the route that a chemical system follows as a reaction proceeds from reactants to products. Locating the transition state in order to determine the SDRP, however, can often be a difficult task. One common approach used to approximate the minimum energy path without a converged transition state structure is to express the pathway as multiple points which are optimized simultaneously[4, 77, 78, 79, 80, 5]. These "chain of states" methods typically begin with a series of images along the linear interpolation between reactants and products, after overall translation and rotation have been removed. These images are updated to minimize the energies of each point subject to constraints, fictitious forces or interpolation/reparameteration schemes, which ensure that the points maintain a uniform description of the pathway. In these methods, the optimizer is generally required to take small steps in order to avoid the introduction of kinks in the path due to the discrete representation of the path. This ad-hoc approach also has the drawback that it is not variational, so there is no reliable way of determining whether or not the optimization is making good progress, or if a solution found is in fact a minimum.

Recently, the line integral of the gradient norm has been described as a variational

property of a reaction path[33, 34]. The line integral of the gradient norm is expressed as

$$\begin{aligned} E_{VRE} &= \int_{t_R}^{t_P} \sqrt{\frac{\partial V\left(\mathbf{x}\left(t\right)\right)}{\partial x}^{\mathrm{T}} \frac{\partial V\left(\mathbf{x}\left(t\right)\right)}{\partial x}} \sqrt{\frac{d\mathbf{x}\left(t\right)}{dt}^{\mathrm{T}} \frac{d\mathbf{x}\left(t\right)}{dt}} dt \\ &= \int_{t_R}^{t_P} \left|\mathbf{g}\left(\mathbf{x}\left(t\right)\right)\right| \left|\boldsymbol{\tau}\left(t\right)\right| dt \end{aligned} \tag{5.1}$$

where V is the potential energy, $\mathbf{x}\left(t\right)$ are the coordinates of the reaction path parameterized by t, t_R and t_P are the parameter values corresponding to the reactant and product structures, respectively, while $\mathbf{g}$ and $\boldsymbol{\tau}$ are used as shorthand for the gradient of the potential, and the tangent to the path. This integral is a non-negative, energetic quantity, and will be referred to as the Variational Reaction Energy (VRE) throughout this work. Rigorous proofs that the VRE is minimized by the steepest descent reaction path are discussed in work by Quapp, Bofill and others, but a simple conceptual proof is provided by computing the VRE assuming that $\mathbf{x}\left(t\right)$ is the steepest descent path. This implies that the tangent to the path is everywhere proportional to the gradient, which simplifies the VRE to the absolute value of the projection of the gradient onto the tangent *(E_{pVRE})*

$$\begin{aligned} E_{pVRE} &= \int_{t_R}^{t_P} \left|\mathbf{g}\left(\mathbf{x}\left(t\right)\right)^{\mathrm{T}} \boldsymbol{\tau}\left(t\right)\right| dt \\ &= \sum_{a} \left(2V\left(\mathbf{x}_{a,TS}\right) - V\left(\mathbf{x}_{a,P}\right) - V\left(\mathbf{x}_{a,R}\right)\right) \end{aligned} \tag{5.2}$$

where a is over the number of barriers along the path, and $\mathbf{x}_{a,R}$ and $\mathbf{x}_{a,P}$ are the local minimum structures adjacent to the corresponding local maximum $\mathbf{x}_{a,TS}$. Since the tangent cannot be changed anywhere along the path without removing the proportionality with the gradient, and since the SDRP steps down from the reaction pathway in the direction that makes the largest change to the magnitude of the gradient, the

SDRP must minimize the VRE. While the projected VRE will only equal the VRE when $\mathbf{x}(t)$ is the SDRP, it may be computed for any arbitrary path connecting the reactants and products. This provides a useful non-negative estimate for the variational error in the current path

$$\begin{aligned} \epsilon &= E_{VRE} - E_{pVRE} \\ &= \int_{t_R}^{t_P} |\mathbf{g}(\mathbf{x}(t))| \, |\boldsymbol{\tau}(t)| \, dt - \sum_a \left(2V(\mathbf{x}_{a,TS}) - V(\mathbf{x}_{a,P}) - V(\mathbf{x}_{a,R})\right) \end{aligned} \tag{5.3}$$

Methods to minimize E_{VRE} by a chain of states approach have been discussed elsewhere, but these suffer from many of the same problems that exist in the ad hoc path optimization methods. Many small steps are required to converge and the discretization error can result in non-variational behavior unless many images are used. Since E_{VRE} is a functional of a smooth, continuous object, it should be advantageous to describe the path using a continuous representation such as a basis set expansion. Such a representation provides a set of coordinates, the linear expansion coefficients (LEC), which can be optimized by minimization of E_{VRE} using standard gradient-based optimization methods. The following work describes the development of a proof of concept for the Variational Reaction Coordinate (VRC) method, and demonstrates the effectiveness of the method on a number of test problems.

5.2 Path Optimization by VRC

5.2.1 A continuous description of the reaction path

To begin outlining the VRC method, a representation of the pathway must be chosen to define $\mathbf{x}(t)$ in terms of a set of parameters that may be varied. A continuous description of the path will remove any ambiguities about how to define the tangent at a given point, or how to choose a distribution of discrete points to provide the

desired level of sampling accuracy between iterations of the optimization. A basis set expansion expresses the path in terms of a set of continuous functions and linear expansion coefficients (LEC, $C_{i\mu}$)

$$\begin{aligned} x_i(t) = \sum_\mu C_{i\mu}\phi_\mu(t) \quad &\rightarrow \quad \frac{\partial x_i(t)}{\partial C_{i\mu}} = \phi_\mu(t) \\ \tau_i(t) = \sum_\mu C_{i\mu}\frac{d\phi_\mu(t)}{dt} \quad &\rightarrow \quad \frac{\partial \tau_i(t)}{\partial C_{i\mu}} = \frac{d\phi_\mu(t)}{dt} \end{aligned} \tag{5.4}$$

where the Roman indices are over the coordinates, and the Greek indices are over the basis functions. The choice of t_R and t_P are completely arbitrary so long as non-zero portions of the basis functions span the space between them, so values of for $t_R = 0$ and $t_P = 1$ will be used throughout this work and functions will be chosen with the appropriate support. Cubic spline functions are commonly used to represent the reaction pathway in discrete path optimization methods in order to produce smooth tangents, or to compute distances between points for constraints or reparameterizations. Therefore, the choice of basis set expansion for the present work will be the B-Splines basis set[81]. B-Splines are a formulation of piecewise continuous polynomial splines, constructed in such a way that they may be expressed as a linear basis set expansion. The B-Spline basis are polynomials, and the number, range, shape and distribution of functions in the basis depends on the choice of the knot vector, $\mathbf{u}$. For the present work, a quartic basis with n LEC per coordinate is defined by n *internal* functions that span the range $0 < t < 1$, along with two *capping* functions which peak sharply at $t = 0$ and $t = 1$. The capping functions allow the geometries of the reactants and products to remain fixed by setting the corresponding LEC to the reactant and product geometries.

The knot vector used in this chapter is defined as follows

$$u_i = \begin{cases} 0 & 1 \leq i \leq d+1 \\ \sum_{j=i-d-1}^{d+1} \frac{j}{n+1} & d+2 \leq i \leq n+2 \\ 1+\delta & n+3 \leq i \leq n+d+3 \end{cases} \tag{5.5}$$

where $d = 4$ is the order of the polynomial, n is the number of internal basis functions, and $\delta = 10^{-10}$ ensures that $\phi(1)$ is defined. The basis functions $\phi_\mu = \phi_\mu^4$ themselves are defined recursively

$$\phi_\mu^0(t) = \begin{cases} 1 & u_\mu \leq t < u_{\mu+1} \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

$$\phi_\mu^k(t) = \frac{t-u_\mu}{u_{\mu+k}-u_\mu}\phi_\mu^{k-1}(t) + \frac{u_{\mu+k+1}-t}{u_{\mu+k+1}-u_{\mu+1}}\phi_{\mu+1}^{k-1}(t) \tag{5.7}$$

with the derivative expressions and more efficient means of evaluating B-splines given in the reference[81]. Figure 5.1 demonstrates the shape of the basis functions when n is 5.

Once a choice of basis has been made, the VRE may be evaluated by an appropriate quadrature method. The determination of what is the most appropriate or efficient method is beyond the scope of the current work, so a simple adaptive quadrature method based upon a combination of 3rd order Gauss-Legendre and 5th order Curtis Clenshaw rules[82] will be used. The adaptive integrator evaluates the integral on a coarse grid as an extrapolation of the two quadrature rules, and computes an error based upon the difference between the two rules. Each interval on the integration grid is evaluated, and any intervals which have an error above a tolerance are evaluated again on a progressively finer grid. This process is repeated until the error for all

intervals is below an absolute or relative tolerance. Once the VRE and it's derivatives have been computed, the energies of all of the PES evaluations used in the adaptive quadrature process can be compared in order to find local maxima/minima along the path for the purpose of computing ϵ_{VRE} and other terms that depend on the location of these stationary points.

5.2.2 VRE derivatives

Having selected a set of coordinates to represent the path, the next step is to construct a local quadratic approximation (LQA, Q_{VRE}) to the VRE

$$\begin{aligned} Q_{VRE}\left(\mathbf{C}+\ \mathbf{C}\right) &= E_{VRE}\left(\mathbf{C}\right)+\Delta\mathbf{C}^{\mathrm{T}}\frac{\partial E_{VRE}\left(\mathbf{C}\right)}{\partial\mathbf{C}}+\frac{1}{2}\Delta\mathbf{C}^{\mathrm{T}}\frac{\partial^{2}E_{VRE}\left(\mathbf{C}\right)}{\partial\mathbf{C}^{2}}\Delta\mathbf{C} \\ &= E_{0}+\Delta\mathbf{C}^{\mathrm{T}}\boldsymbol{\gamma}_{0}+\frac{1}{2}\Delta\mathbf{C}^{\mathrm{T}}\boldsymbol{\eta}_{0}\Delta\mathbf{C} \end{aligned} \tag{5.8}$$

Where γ and η are used to represent the gradient and Hessian of the VRE with respect to a change in the LEC in order to avoid confusion with the gradient and Hessian of the potential with respect to a change in the molecular geometry, represented by $\mathbf{g}$ and $\mathbf{H}$, respectively. The formula for the VRE gradient is given below, with the explicit dependence on $\mathbf{x}$ and t dropped for brevity

$$\begin{aligned} \gamma_{i\mu}=\frac{\partial E_{VRE}}{\partial C_{i\mu}} &= \left|\mathbf{g}_{P}\right|\left|\boldsymbol{\tau}_{P}\right|\frac{\partial\mathbf{x}_{P}}{\partial C_{i\mu}}-\left|\mathbf{g}_{R}\right|\left|\boldsymbol{\tau}_{R}\right|\frac{\partial\mathbf{x}_{R}}{\partial C_{i\mu}}+\int_{t_{R}}^{t_{P}}\frac{\partial\left(\left|\mathbf{g}\right|\left|\boldsymbol{\tau}\right|\right)}{\partial C_{i\mu}}dt \\ &= \int_{t_{R}}^{t_{P}}\left(\frac{\partial\left|\mathbf{g}\right|}{\partial C_{i\mu}}\left|\boldsymbol{\tau}\right|+\left|\mathbf{g}\right|\frac{\partial\left|\boldsymbol{\tau}\right|}{\partial C_{i\mu}}\right)dt \end{aligned} \tag{5.9}$$

The two terms outside of the integral can be safely neglected since the coefficients corresponding to the capping functions are fixed, and the reactants and products do not vary with changes to the internal functions. Differentiating equation 5.9 a second

time yields

$$\eta_{i\mu j\nu} = \frac{\partial^2 E_{VRE}}{\partial C_{i\mu}\partial C_{j\nu}} = \int_{t_R}^{t_P} \left(\frac{\partial^2 |\mathbf{g}|}{\partial C_{i\mu}\partial C_{j\nu}} |\boldsymbol{\tau}| + \frac{\partial |\mathbf{g}|}{\partial C_{i\mu}} \frac{\partial |\boldsymbol{\tau}|}{\partial C_{j\nu}} + \frac{\partial |\mathbf{g}|}{\partial C_{j\nu}} \frac{\partial |\boldsymbol{\tau}|}{\partial C_{i\mu}} + |\mathbf{g}| \frac{\partial^2 |\boldsymbol{\tau}|}{\partial C_{i\mu}\partial C_{j\nu}} \right) dt \tag{5.10}$$

The derivatives of the gradient norm and the tangent norm with respect to changes in LEC are straightforward to compute

$$\frac{\partial |\mathbf{g}|}{\partial C_{i\mu}} = \frac{\sum_a H_{ia} g_a}{|\mathbf{g}|} \phi_\mu \tag{5.11}$$

$$\frac{\partial |\boldsymbol{\tau}|}{\partial C_{i\mu}} = \frac{\tau_i}{|\boldsymbol{\tau}|} \frac{d\phi_\mu}{dt} \tag{5.12}$$

$$\frac{\partial^2 |\mathbf{g}|}{\partial C_{i\mu}\partial C_{j\nu}} = \left(\frac{\sum_a \left(\partial H_{ia}/\partial x_j\right) g_a + H_{ia} H_{ja}}{|\mathbf{g}|} - \frac{\sum_{a,b} H_{ia} g_a H_{jb} g_b}{|\mathbf{g}|^3} \right) \phi_\mu \phi_\nu \tag{5.13}$$

$$\frac{\partial^2 |\boldsymbol{\tau}|}{\partial C_{i\mu}\partial C_{j\nu}} = \left(\frac{\delta_{ij}}{|\boldsymbol{\tau}|} - \frac{\tau_i \tau_j}{|\boldsymbol{\tau}|^3} \right) \frac{d\phi_\mu}{dt} \frac{d\phi_\nu}{dt} \tag{5.14}$$

Since the VRE depends on the potential energy gradient, the VRE gradient depends on the potential energy Hessian, and the VRE Hessian depends on the third derivative of the potential energy. However, in each of these cases, it is only the product of the higher derivative with the gradient that is necessary, which may be computed numerically by reverse difference.

$$(Tg)_{ij} = \sum_a \frac{\partial H_{ij}}{\partial x_a} g_a \approx \frac{|\mathbf{g}|}{\delta} \left(H_{ij}(\mathbf{x}) - H_{ij}\left(\mathbf{x} - \delta \frac{\mathbf{g}}{|\mathbf{g}|} \right) \right) \tag{5.15}$$

Combining equations 5.9 and 5.10 with equations 5.11-5.15, the full expressions for

the VRE gradient and Hessian are

$$\gamma_{i\mu} = \int_{t_R}^{t_P} \left(\frac{|\boldsymbol{\tau}|}{|\mathbf{g}|} \sum_a H_{ia} g_a \phi_\mu + \frac{|\mathbf{g}|}{|\boldsymbol{\tau}|} \tau_i \frac{d\phi_\mu}{dt} \right) dt \tag{5.16}$$

$$\begin{aligned} \eta_{i\mu j\nu} = \int_{t_R}^{t_P} \Bigg(& \frac{|\boldsymbol{\tau}|}{|\mathbf{g}|} \left((Tg + HH)_{ij} - \frac{\sum_{a,b} H_{ia} g_a H_{jb} g_b}{|\mathbf{g}|^2} \right) \phi_\mu \phi_\nu + \\ & \frac{\sum_a H_{ia} g_a \tau_i}{|\mathbf{g}|\,|\boldsymbol{\tau}|} \phi_\mu \frac{d\phi_\nu}{dt} + \frac{\tau_i \sum_a H_{ja} g_a}{|\mathbf{g}|\,|\boldsymbol{\tau}|} \frac{d\phi_\mu}{dt} \phi_\nu + \\ & |\mathbf{g}| \left(\frac{\delta_{ij}}{|\boldsymbol{\tau}|} - \frac{\tau_i \tau_j}{|\boldsymbol{\tau}|^3} \right) \frac{d\phi_\mu}{dt} \frac{d\phi_\nu}{dt} \Bigg) dt \end{aligned} \tag{5.17}$$

With the VRE gradient and Hessian computed, Newton's method can be used to find the LEC displacement corresponding to the minimum of a shifted VRE LQA (Eqn. 5.8)

$$\boldsymbol{\Delta C} = -\left(\boldsymbol{\eta} - \lambda_\sigma \boldsymbol{\sigma}\right)^{-1} \boldsymbol{\gamma} \tag{5.18}$$

where $\boldsymbol{\sigma}$ is a positive definite shift matrix and λ_σ is chosen such that the shifted Hessian $(\boldsymbol{\eta} - \lambda_\sigma \boldsymbol{\sigma})$ is positive definite and the step size is reasonable. In geometry optimizations, the shift matrix is often taken to be the identity matrix, and the rational function optimization method is used to compute λ_σ . In some problems, such as those with strongly coupled coordinates or ill-conditioned Hessians, the use of the identity matrix can lead to numerical instabilities or slow convergence. Over the course of testing and implementing the VRC method, the overlap of the basis set derivatives was found to provide better optimization behavior than the identity matrix

$$\sigma_{i\mu j\nu} = \int_{t_R}^{t_P} \delta_{ij} \frac{d\phi_\mu}{dt} \frac{d\phi_\nu}{dt} dt \tag{5.19}$$

Standard methods for computing an optimization step with the Hessian shifted by

the identity matrix, such as the rational function optimization (RFO) method[54], may be used after scaling eqn 5.18 as follows

$$\mathbf{\Delta C} = -\boldsymbol{\sigma}^{-1/2}\left(\boldsymbol{\sigma}^{-1/2}\boldsymbol{\eta}\boldsymbol{\sigma}^{-1/2} - \lambda_\sigma \mathbf{I}\right)^{-1}\boldsymbol{\sigma}^{-1/2}\boldsymbol{\gamma} \tag{5.20}$$

In addition to controlling the step size and direction, λ_σ can also be adjusted to ensure that the predicted error following the step is greater than or equal to zero. A quadratic approximation to the current error, $Q_\epsilon\left(\mathbf{C} + \mathbf{\Delta C}\right)$ with respect to a change in the LEC can be constructed by differentiating equation 5.3. The unconstrained VRC optimization is as follows

1. Input initial path, maximum step size δ
2. Compute VRE, VRE derivatives, ϵ and $\boldsymbol{\sigma}$.
3. Set λ_σ to the RFO eigenvalue using Hessian/gradient scaled by $\boldsymbol{\sigma}^{-1/2}$ as in eq 5.20
4. Compute displacement $\mathbf{\Delta C}$ by eq 5.18, if $|\mathbf{\Delta C}| \geq \delta$, decrease λ_σ until $|\mathbf{\Delta C}| = \delta$
5. Compute Q_ϵ, if $Q_\epsilon \leq 0$, decrease λ_σ until $Q_\epsilon = 0$.
6. Check $|\mathbf{\Delta C}|$ for convergence, stop iterations if converged
7. Update LEC, recompute VRE, VRE derivatives, ϵ and $\boldsymbol{\sigma}$.
8. Compare predicted change in energy to actual change in energy, and update δ accordingly.
9. Goto 3

This algorithm is capable of producing final pathways with very little error, even with a small number of LEC per coordinate. Throughout the optimization, steady progress is made in the direction of the final pathway, however much of the improvement in the path appears to take place in the early steps, and there is a large number of steps where the shape of the path, as well as the VRE and the magnitude of the VRE

gradient, do not appear to change by much until the last few optimization steps where the behavior of the optimization appears to exhibit quadratic convergence. This sort of optimization behavior suggests strongly that there are degrees of freedom in the LEC for which the VRE is invariant, and the algorithm needs to be modified to account for these degrees of freedom.

5.2.3 Constraints and constrained optimization

For a single pathway, there may be more than one set of LEC that closely describes the shape of a particular path in a given basis set. Since the VRE is a line integral which is invariant to the chosen representation of the pathway, these two sets of LEC will have the same energy, and the value of both the first and the second derivative of the VRE in the direction of the displacement from one set to the other will be zero. In an ideal optimization utilizing an infinite basis set and computing the VRE integrals exactly, these redundant coordinates would be pure and separable and could be identified and eliminated at each iteration in order to accelerate and stabilize the convergence to the SDRP in much the same way as translation and rotation are removed from optimization of single geometries. With a finite basis set and numerical quadrature methods, however, such pure transformations do not necessarily exist, and the coupling between the redundant and non-redundant LEC can complicate the removal of the redundant coordinates from consideration.

In order to develop a more robust and reliable means to deal with the redundant coordinates in the LEC, it's useful to begin by discussing what gives rise to them through discussion of curve fitting. If one were interested in finding a path $\mathbf{x}(t)$ expressed as an expansion of n basis functions such that the path passed through a set of n geometries $\mathbf{x}_i$, the typical strategy would be to solve a set of $n\ x\ m$ equations $\mathbf{x}(t_i) = \mathbf{x}_i$ for the $n\ x\ m$ LEC, where m is the dimensionality of the geometries $\mathbf{x}_i$. These t_i (or a process

by which to obtain them numerically from the $\mathbf{x}_i$) must be defined in order to evaluate the basis functions and carry out the fit, and the LEC computed by solving the linear equations is unique for different values of t_i. More directly, the LEC corresponding to the path fit through the geometries is a function of both the geometries themselves, as well as the distance (in t) between the points along the path; the representation of a path embedded in an m-dimensional space and defined by n basis functions requires $n\ x\ (m+1)$ values to be defined uniquely. Figure 5.2 illustrates this with a simple example, three cubic spline curves are fit to the same 5 points, using different values of t for the 3 internal points. The result is 3 curves with quite different properties.

The VRC depends on the shape of the path, but not on the relative spacing between the points along the path. Ideally, the constraints would have the opposite relationship, depending on the relative spacing of points along the path while being as independent of the shape as possible. Even though the VRE derivatives and LEC displacements are expressed in terms of continuous quantities rather than discrete points along the path, the analogy to curve fitting suggests that the best way to find a unique set of LEC for n basis functions that approximates a path with a particular shape is to define the relationship between n values of t along the path. This can be done by using the arc length, S, which is a convenient and unambiguous way to define the distance between two points that does not otherwise depend on the shape of the path:

$$S(t_1, t_2) = \int_{t_1}^{t_2} |\boldsymbol{\tau}(t)|\, dt \tag{5.21}$$

For n basis functions, the relationship between the parameter t and the arc length must be defined for n values of t. This can be most easily accomplished by dividing the path into $n+1$ segments, and requiring that the ratio of the lengths of adjacent

segments is equal to some constant

$$t_a = \frac{\alpha}{n+1} \qquad 0 \leq a \leq n+1 \tag{5.22}$$

$$S(t_{a-1}, t_a) = c_a S(t_a, t_{a+1}) \qquad 1 \leq a \leq n+1 \tag{5.23}$$

where the constants c_i determine how much of the basis set is dedicated to modeling a given segment. By making the c_i depend on properties of the path or PES the flexibility of the basis set could be focused on the regions of large curvature or high relative energies that may be the most important for understanding the reaction. In the present work, however, the c_i are always chosen to be equal to 1 in order to maintain a uniform description of the path. This leads to n constraint functions

$$\kappa_a = 0 = S(t_{a-1}, t_a) - S(t_a, t_{a+1}) \qquad 1 \leq a \leq n \tag{5.24}$$

The method of Lagrange multipliers may be used to enforce these constraints during the minimization of the VRE

$$\mathcal{L}_{CVRC} = Q_{VRE}\left(\mathbf{C} + \mathbf{\Delta C}\right) - \frac{1}{2}\lambda_\sigma \Delta \mathbf{C}^{\mathrm{T}} \boldsymbol{\sigma}\left(\mathbf{C}\right) \Delta \mathbf{C} + \sum_\alpha \lambda_\alpha \kappa_\alpha \left(\mathbf{C} + \mathbf{\Delta C}\right) \tag{5.25}$$

Equation 5.25 is the constrained VRC (CVRC) Lagrangian, where λ_σ is the shift parameter chosen to ensure that a downhill step is taken that was used in the unconstrained method, and is computed using the same scaled RFO approach as before. The path may be updated iteratively towards the solution by requiring that $\mathcal{L}_{CVRC}$ is stationary with respect to a change in both the LEC ($\mathbf{\Delta C}$) and the multipliers λ_α

$$\frac{\partial \mathcal{L}}{\partial C} = 0 \text{ and } \frac{\partial \mathcal{L}}{\partial \lambda_\alpha} = \kappa_\alpha = 0 \tag{5.26}$$

The integrals required to compute the terms in Q_{VRE} depend on the potential energy of the surface and therefore are computationally expensive to compute, while the integrals necessary to compute κ_α and it's derivatives with respect to a change in the LEC depend only on the evaluation of the basis functions and are relatively inexpensive to compute. Since both Q_{VRE} and the κ_α have a strongly curvilinear dependence upon the LEC, it makes sense to solve for the $\mathbf{\Delta C}$ and λ_α using a microiterative approach for each evaluation of the VRE derivatives $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$. Although $\boldsymbol{\sigma}$ depends only on the basis functions, the role it serves is as a shifting matrix for $\boldsymbol{\eta}$, and therefore, is only recomputed whenever $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are evaluated.

In addition to minimizing the VRE under the constraint that n points along the path remain uniformly spaced, it would be advantageous to also include terms to control step size and restrict the solution to displacements with predicted VRE error greater than zero. In order to do this, the following additional terms can be added to equation 5.25:

$$\frac{1}{2}\mu_\delta \left(\mathbf{\Delta C}^T \mathbf{\Delta C} - \delta^2\right) \tag{5.27}$$

$$\mu_\epsilon \boldsymbol{\epsilon} \left(\mathbf{C} + \mathbf{\Delta C}\right) \tag{5.28}$$

where the μ are multipliers, δ is the maximum step size, and ϵ is the error as defined by equation 5.3, expanded as a quadratic Taylor series in $\mathbf{\Delta C}$ with the derivatives given by:

$$\frac{\partial \epsilon}{\partial C_{i\mu}} = \gamma_{i\mu} - 2g_i\left(\mathbf{x}\left(t_{ts}\right)\right)\phi_\mu\left(t_{ts}\right) \tag{5.29}$$

$$\frac{\partial^2 \epsilon}{\partial C_{i\mu}\partial C_{j\nu}} = \eta_{i\mu j\nu} - 2H_{ij}\left(\mathbf{x}\left(t_{ts}\right)\right)\phi_\mu\left(t_{ts}\right)\phi_\nu\left(t_{ts}\right) \tag{5.30}$$

The first term is not included in the Lagrangian unless the microiterations pro-

duce an LEC displacement with a magnitude that exceeds δ, while the second term is only included when the predicted error falls below zero. The derivatives of $\mathcal{L}_{CVRC}$ during each phase of the microiterations can be computed by expanding $\mathbf{\Delta C}$ and the multipliers with linear displacements:

$$\mathcal{L}_{CVRC}\left(\mathbf{C}+\mathbf{\Delta C},\boldsymbol{\lambda}_{\kappa},\mu_{\delta},\mu_{\epsilon}\right)=\mathcal{L}_{CVRC}\Big(\mathbf{C}+\mathbf{\Delta C}+\mathbf{\Delta\Delta C},\boldsymbol{\lambda}_{\kappa}+\mathbf{\Delta}\boldsymbol{\lambda}_{\kappa},\\ \mu_{\delta}+\Delta\mu_{\delta},\mu_{\epsilon}+\Delta\mu_{\epsilon}\Big) \tag{5.31}$$

$$\frac{\partial\mathcal{L}_{CVRC}}{\partial\mathbf{\Delta C}}=\boldsymbol{\gamma}+\left(\boldsymbol{\eta}-\lambda_{\sigma}\boldsymbol{\sigma}+\mu_{\delta}\mathbf{I}\right)\mathbf{\Delta C}+\sum_{\alpha}\lambda_{\alpha}\frac{\partial\kappa_{\alpha}\left(\mathbf{C}+\mathbf{\Delta C}\right)}{\partial\mathbf{C}}+\\ \mu_{\epsilon}\frac{\partial\epsilon_{VRC}\left(\mathbf{C}+\mathbf{\Delta C}\right)}{\partial\mathbf{C}} \tag{5.32}$$

$$\frac{\partial\mathcal{L}_{CVRC}}{\partial\lambda_{\alpha}}=\kappa_{\alpha} \tag{5.33}$$

$$\frac{\partial\mathcal{L}_{CVRC}}{\partial\mu_{\delta}}=\frac{1}{2}\left(\mathbf{\Delta C}^{T}\mathbf{\Delta C}-\delta^{2}\right) \tag{5.34}$$

$$\frac{\partial\mathcal{L}_{CVRC}}{\partial\mu_{\epsilon}}=\epsilon\left(\mathbf{C}+\mathbf{\Delta C}\right) \tag{5.35}$$

$$\frac{\partial^{2}\mathcal{L}_{cVRC}}{\partial\mathbf{\Delta C}^{2}}=\boldsymbol{\eta}-\lambda_{\sigma}\boldsymbol{\sigma}+\mu_{\delta}\mathbf{I}+\sum_{\alpha}\lambda_{\alpha}\frac{\partial^{2}\kappa_{\alpha}\left(\mathbf{C}+\mathbf{\Delta C}\right)}{\partial\mathbf{C}^{2}}+\\ \mu_{\epsilon}\frac{\partial^{2}\epsilon_{VRC}\left(\mathbf{C}+\mathbf{\Delta C}\right)}{\partial\mathbf{C}^{2}} \tag{5.36}$$

$$\frac{\partial^{2}\mathcal{L}_{cVRC}}{\partial\mathbf{\Delta C}\partial\lambda_{\alpha}}=\frac{\partial\kappa_{\alpha}\left(\mathbf{C}+\mathbf{\Delta C}\right)}{\partial\mathbf{C}} \tag{5.37}$$

$$\frac{\partial^{2}\mathcal{L}_{cVRC}}{\partial\mathbf{\Delta C}\partial\mu_{\delta}}=\mathbf{\Delta C} \tag{5.38}$$

$$\frac{\partial^{2}\mathcal{L}_{cVRC}}{\partial\mathbf{\Delta C}\partial\mu_{\epsilon}}=\frac{\partial\epsilon_{VRC}\left(\mathbf{C}+\mathbf{\Delta C}\right)}{\partial\mathbf{C}} \tag{5.39}$$

$$\begin{pmatrix} \boldsymbol{\Delta\Delta C} \\ \boldsymbol{\Delta\lambda}_\kappa \\ \boldsymbol{\Delta\mu}_\delta \\ \boldsymbol{\Delta\mu}_\epsilon \end{pmatrix}^T = - \begin{pmatrix} \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\Delta C}^2} & \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\Delta C}\partial\boldsymbol{\lambda}_\kappa}^T & \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\Delta C}\partial\mu_\delta}^T & \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\Delta C}\partial\mu_\epsilon}^T \\ \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\Delta C}\partial\boldsymbol{\lambda}_\kappa} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\Delta C}\partial\mu_\delta} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\Delta C}\partial\mu_\epsilon} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial\mathcal{L}}{\partial\boldsymbol{\Delta C}} \\ \frac{\partial\mathcal{L}}{\partial\boldsymbol{\lambda}_\kappa} \\ \frac{\partial\mathcal{L}}{\partial\mu_\delta} \\ \frac{\partial\mathcal{L}}{\partial\mu_\epsilon} \end{pmatrix} \tag{5.40}$$

The CVRC algorithm is as follows

1. Input initial path, maximum step size δ
2. Compute VRE, VRE derivatives, ϵ and $\boldsymbol{\sigma}$
3. Set λ_σ to the RFO eigenvalue using Hessian/gradient scaled by $\boldsymbol{\sigma}^{-1/2}$ as in eq 5.20, set $\boldsymbol{\Delta C}$, $\boldsymbol{\lambda}_\kappa$, μ_δ and μ_ϵ to $\mathbf{0}$
4. Begin microiterations
   (a) Compute the constraints $\kappa(\mathbf{C} + \boldsymbol{\Delta C})$ and their derivatives w.r.t. a change in the LEC
   (b) Compute ϵ and $|\boldsymbol{\Delta C}|$ and turn on optimization of μ_δ and μ_ϵ if necessary
   (c) Compute derivatives of the Lagrangian according to eqs 5.32-5.39
   (d) Update $\boldsymbol{\Delta C}$, $\boldsymbol{\lambda}_\kappa$, μ_δ and μ_ϵ according to eq 5.40
   (e) Check augmented gradient and augmented displacement for convergence of microiterations, end microiterations if converged
   (f) goto 4a
5. Check $|\boldsymbol{\Delta C}|$ for convergence, and end macroiterations if converged
6. Update LEC for path, recompute VRE, VRE derivatives, ϵ and $\boldsymbol{\sigma}$
7. Compare predicted versus actual change in VRE, and update δ accordingly
8. goto 3

The constrained VRC algorithm not only manages to get close to the IRC path in fewer iterations than the unconstrained algorithm, it also manages to achieve full convergence much more quickly. The primary drawback to using the constraints is that

the flexibility in the LEC is reduced in order to satisfy the constraints, which results in a higher VRE at the final converged path than in the unconstrained case. Additionally, λ_σ does not approach zero near convergence as it does in the unconstrained case. This is likely because it is computed using the unconstrained η which may not be positive definite, and the unconstrained γ which may be non-zero in the direction of the constraints. Early attempts to consider the constraints in the calculation of λ_σ or to include the optimization of λ_σ in the microiteration phase resulted in a loss of stability in the algorithm. It is possible that convergence may be accelerated near the solution by improved handling/computation of λ_σ, and so future investigation is warranted.

Both the constrained and unconstrained algorithms have a tendency to slow down or produce poor step directions early on, when the path is in a region of the PES with incorrect curvature. This is an unfortunate consequence of the VRE's dependence on the gradient norm, as the gradient norm will also be small near higher order stationary points on the PES. This can also complicate the calculation of the VRE Hessian, since eq 5.13 becomes singular when the PES gradient goes to zero. These features can result in steps that are unnecessarily small or cautious as the VRC method has a strong preference to avoid an increase in the gradient norm along the path even early in an optimization where it may be more sensible to focus on reducing the energy of the transition state. In the following section, a modification to the CVRC algorithm is outlined that couples together a standard transition state optimization step with the VRC path relaxation in order to improve the efficiency of the method when the path is far from convergence.

5.3 Combined path and transition state optimization

5.3.1 Coupling constraints

Path optimization methods are often used to produce an approximate geometry corresponding to the transition state connecting two minimum energy structures, which is then further refined by saddle point optimization methods. Path optimization typically requires a significant number of potential energy surface evaluations to produce an approximate structure, but the resulting approximate structures tend to converge more rapidly and/or reliably to the transition state than simpler interpolation schemes like LST/QST[83, 32], or local optimization methods like the dimer method[59]. Additionally, the converged path is usually sufficient to confirm that the transition state does connect the minimum energy structures, so further improvement of the path by reaction path following is not necessary. In existing discrete path optimization methods, the approximation of a transition state geometry is usually accomplished either by interpolation between the highest energy structures following convergence of the path, or by treating the highest energy structure (typically called the *climbing* structure) differently than the rest in order to allow it to loosely converge to the saddle point rather than an arbitrary point near the SDRP.

The VRC method expresses and optimizes the path as a single, continuous object, so producing a geometry to refine to the transition state following the VRC optimization would be a fairly trivial and straightforward optimization of the potential energy with respect to the parameter t. The second approach is less straightforward to adapt to the VRC method, and before discussing how this can be accomplished, it is worthwhile to consider what effect such a coupling would have on a continuous description of the path. In a similar fashion to how a discrete path optimization assigns one point along the path to be a climbing structure, the coupling of a transition state optimization

to the VRC method could be thought of as assigning one basis functions worth of the path's flexibility to the ts optimization. The path for a chemical system described by m coordinates and expanded in a basis of n functions has m x n degrees of freedom minus the n constraints described earlier. Requiring that the path passes through a particular point (i.e. $\mathbf{x}(t_{ts}) = \mathbf{x}_{ts}$) amounts to setting an additional m coupling constraints:

$$0 = \Delta_i x_{ts} = x_i(t_{ts}) - x_{ts,i} \qquad 1 \leq i \leq m \tag{5.41}$$

while introducing an extra degree of freedom in t_{ts}. The inclusion of t_{ts} as an extra degree of freedom highlights a significant benefit to using a continuous description of the path: the location of the transition state along the path is entirely independent of the representation of the path and the evaluation of the VRE or it's derivatives, as well as the constraints from the previous section that define the relationship between the arc length and t.

As with the arc length constraints used in the previous section, the transition state coupling constraints will be enforced during the optimization through the use of Lagrange multipliers which will be determined microiteratively. Each coupling constraint has two terms, the first of which, $x_i\left(t_{ts}\right)$, is the evaluation of coordinate i at the parameter value t_{ts} and therefore depends on the current value of the LEC during the microiterations (i.e. $\mathbf{C} + \mathbf{\Delta C}$). The second term $x_{ts,i}$ is the goal value of coordinate i for this transition state at this iteration of the optimization. This goal structure could be defined implicitly as a functional of the LEC, for example by using the predicted PES gradient or energy at $\mathbf{x}\left(t_{ts}\right)$, which would allow the goal structure to be updated during the microiterations. This approach could have some benefit, but the present discussion will be limited to an explicit definition of $\mathbf{x}_{ts}$, where the goal geometry is

computed once per macroiteration using a modified Newton step on the PES from a point along the path for the current macroiteration and is considered to be fixed during the microiterative portion of the algorithm. This separation allows for the use of standard methods like step size control and line search on the more familiar chemical PES, rather than the VRE potential, and allows the transition state optimization to be viewed as a means to focus the VRC optimization toward a particular region of the PES that's more likely to contain the transition state, and therefore, the SDRP. For this reason, this approach will be referred to as the focused VRC (FVRC) method for the remainder of this paper.

The method for computing the goal structure is discussed in greater detail in section 5.3.2, but for now, let $\mathbf{\Delta x}$ be the m-dimensional array of coupling constraints defined as in eq 5.41. The FVRC Lagrangian is given below

$$\begin{aligned}\mathcal{L}_{FVRC} =& Q_{VRE}\left(\mathbf{C}+\mathbf{\Delta C}\right) - \frac{1}{2}\lambda_\sigma \Delta\mathbf{C}^{\mathrm{T}}\boldsymbol{\sigma}\left(\mathbf{C}\right)\Delta\mathbf{C} + \mu_\epsilon \epsilon_{FVRC}\left(\mathbf{C}+\mathbf{\Delta C}\right) \\ &+ \sum_\alpha \lambda_\alpha \kappa_\alpha \left(\mathbf{C}+\mathbf{\Delta C}\right) + \left(\mathbf{y}+\frac{1}{2}\mathbf{\Delta x}\right)^T \mathbf{\Delta x} \end{aligned} \tag{5.42}$$

where the $\mathbf{y}$ are the multipliers for the coupling constraints. Aside from the addition of the coupling constraint term, there are changes to the step size and error terms compared to $\mathcal{L}_{CVRC}$. The step size term is dropped entirely since it may lead to an inconsistent Lagrangian if the LEC step size is too small to satisfy the coupling constraints, and is unnecessary since the transition state optimization has a controlled step size and limiting the step size of the transition state is sufficient to limit the change in the LEC. Recall that the error is defined as the VRE minus the projected VRE, and that the projected VRE is a sum of the forward and reverse energy barriers. So long as the coupling constraints are satisfied, the estimated projected VRE does not depend

on the LEC and the FVRC error simplifies to

$$\epsilon_{FVRC}\left(\mathbf{C}+\mathbf{\Delta C}\right)=Q_{VRE}\left(\mathbf{C}+\mathbf{\Delta C}\right)-E_{ppVRE} \tag{5.43}$$

where E_{ppVRE} is the predicted projected VRE, evaluated using computed or estimated energies corresponding to the geometry updated by the transition state optimization step. Since E_{ppVRE} is constant with respect to a change in the LEC, its derivatives are equal to the derivatives of the VRE LQA. This approximation is only valid when the path passes through the updated geometry, so optimization of μ_ϵ should not be attempted unless the predicted error is less than zero *and* $\mathbf{\Delta x}=0$. Another consequence of defining the error as being relative to the projected VRE of the final path is that the only term in $\mathcal{L}_{FVRC}$ that depends on t_{ts} is the coupling constraint term. As a notational convenience, let the coupling constraint term be $\mathcal{F}=\left(\mathbf{y}+\frac{1}{2}\mathbf{\Delta x}\right)^T\mathbf{\Delta x}$. The derivatives of $\mathcal{F}$ with respect to a change in the LEC, the coupling constraint multipliers, and t_{ts} are derived straightforwardly:

$$\frac{\partial\mathcal{F}}{\partial y_i}=\Delta_i x_{ts} \tag{5.44}$$

$$\frac{d\mathcal{F}}{dt_{ts}}=\left(\mathbf{y}+\mathbf{\Delta x}\right)^T\boldsymbol{\tau}\left(t_{ts}\right) \tag{5.45}$$

$$\frac{\partial\mathcal{F}}{\partial C_{i\mu}}=\left(y_i+\Delta_i x_{ts}\right)\phi_\mu\left(t_{ts}\right) \tag{5.46}$$

$$\frac{d\partial\mathcal{F}}{dt_{ts}\partial y_i}=\tau_i\left(t_{ts}\right) \tag{5.47}$$

$$\frac{d^2\mathcal{F}}{dt_{ts}^2}=\left(\mathbf{y}+\mathbf{\Delta x}\right)^T\frac{d\boldsymbol{\tau}\left(t_{ts}\right)}{dt}+\boldsymbol{\tau}\left(t_{ts}\right)^T\boldsymbol{\tau}\left(t_{ts}\right) \tag{5.48}$$

$$\frac{d\partial\mathcal{F}}{dt_{ts}\partial C_{i\mu}}=\tau_i\left(t_{ts}\right)\phi_\mu\left(t_{ts}\right)+\left(y_i+\Delta_i x_{ts}\right)\frac{d\phi_\mu\left(t_{ts}\right)}{dt} \tag{5.49}$$

$$\frac{\partial^2\mathcal{F}}{\partial C_{i\mu}\partial y_j}=\delta_{ij}\phi_\mu\left(t_{ts}\right) \tag{5.50}$$

$$\frac{\partial^2\mathcal{F}}{\partial C_{i\mu}\partial C_{j\nu}}=\delta_{ij}\phi_\mu\left(t_{ts}\right)\phi_\nu\left(t_{ts}\right) \tag{5.51}$$

5.3.2 Geometry optimization

As mentioned earlier, the difficulty in transition state optimization is a result of the requirement that the energy must be a maximum along the transition vector, while being a minimum in all other degrees of freedom. Not only is the selection of the transition vector difficult when far from the converged structure, but methods, like line searches, which are commonly used to accelerate the convergence of minimum energy structures cannot be used as there is no suitable metric to search. Since the energy must be a maximum in one direction, the step length that minimizes the energy may not be optimal, and since the magnitude of the gradient needs to increase if the curvature of the Hessian is incorrect, a step length that minimizes the gradient norm or gradient squared may not be optimal. One additional benefit of the focused VRC method is that it turns the difficult problem of computing a step towards a transition state into the much simpler problem of computing a step that minimizes the energy.

Since t_{ts} corresponds to the maximum along the current path, and since the maximum along the current path must be greater than or equal to the energy of the converged path, there's no need to select an eigenvector to be the transition vector when computing $\mathbf{x}_{ts}$. If the PES Hessian $\mathbf{H}_0$ at $\mathbf{x}_0 = \mathbf{x}(t_{ts})$ has the correct number of negative eigenvalues (one for a transition state), the corresponding eigenvector must be the transition vector and Newton steps should suffice to converge to the transition state:

$$\mathbf{x}_{TS} = \mathbf{x}_0 - a_{scl}\mathbf{H}_0^{-1}\mathbf{g}_0 \tag{5.52}$$

where $\mathbf{g}_0$ is the gradient at $\mathbf{x}_0$ and a_{scl} is a scale factor which will be discussed later. If $\mathbf{H}_0$ has more than one negative eigenvalue or produces a step larger than the allowed step size with a_{scl}, a downhill step orthogonal to the tangent $\boldsymbol{\tau}$ of current path, is used

instead:

$$\mathbf{P}_{\tau}^{\parallel} = \frac{\boldsymbol{\tau}\boldsymbol{\tau}^T}{\boldsymbol{\tau}^T\boldsymbol{\tau}}; \qquad \mathbf{P}_{\tau}^{\perp} = \mathbf{I} - \mathbf{P}_{\tau}^{\parallel} \tag{5.53}$$

$$\mathbf{x}_{TS} = \mathbf{x}_0 - a_{scl}\left(\mathbf{P}_{\tau}^{\perp}\mathbf{H}_0\mathbf{P}_{\tau}^{\perp} - \lambda_{rfo}\mathbf{I} + \mathbf{P}_{\tau}^{\parallel}\right)^{-1}\mathbf{P}_{\tau}^{\perp}\mathbf{g}_0 \tag{5.54}$$

Since the step must lower the PES energy, regardless of how it's computed, a line search may be employed to improve the quality of the goal structure. To carry out the line search, an $\mathbf{x}_1$ is computed according to eqn 5.54 or eqn 5.52 with $a_{scl} = a_1 \leq 1$ set so that $|\Delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_0| \leq \delta_{max}$, where δ_{max} is the maximum allowed stepsize for the transition state. A fourth-order polynomial

$$p(\alpha) = c_0 + c_1\alpha + c_2\alpha^2 + c_3\alpha^3 + c_4\alpha^4 \tag{5.55}$$

can be constructed by using the energies $p(0) = V_0$ and $p(1) = V_1$ and the scalar gradients $p'(0) = \mathbf{g}_0^T\Delta\mathbf{x}$ and $p'(1) = \mathbf{g}_1^T\Delta\mathbf{x}$ as well as a constraint that there is only one minimum ($p''(\alpha) \geq 0$). If such a polynomial does not exist, then a quartic polynomial with a zero cubic term $c_3 = 0$ is constructed instead. The minimum α_{min} closest to $\alpha = 1$ is located, and $\mathbf{x}_{TS}$ is updated using $a_{scl} = a_1\alpha_{min}$.

5.3.3 Handling rotations

One of the more attractive features of the VRC method is that no extra considerations need to be made for handling overall translation or rotation when working with cartesian coordinates. By definition, an infinitesimal translation or rotation of the geometry along the path will not change the magnitude of the gradient of the internal energy of a molecule, but it may increase the overall distance the path needs to travel from reactant to product. Because of this, minimization of the VRE will also minimize the overall translation and rotation contained in the path. Unfortunately, the same cannot be said about the coupling constraints defined in eqn 5.41.

Translation may be easily accounted for by requiring that the reactant and product both be *mean centered* by translating the atoms so that the average position of the x, y and z coordinate of the atoms in each structure are all zero. Rotation is a bit more difficult, as the goal structure is always defined in a particular rotational orientation, which may not necessarily be the optimal rotational orientation to minimize the VRE of a path that passes through the internal coordinates for that point. Extra care must be taken to limit the impact that the rotational orientation of the goal structure has on the relaxation of the path.

A projection method can be used to eliminate any translation or rotation from the geometry optimization step

$$\mathbf{P}_{TR} = \sum_{i}^{3} \frac{1}{n_{atoms}} \mathbf{t}_i \mathbf{t}_i^T + \mathbf{r}_i \left(\sum_{j}^{3} \mathbf{r}_j^T \mathbf{r}_j \right)^{-1} \mathbf{r}_i^T \tag{5.56}$$

where the portion of these vectors corresponding to the kth atom are given by

$$\mathbf{t}_{i,k} = \mathbf{e}_i \tag{5.57}$$

$$\mathbf{r}_{i,k} = \mathbf{x}_k \times \mathbf{e}_i \tag{5.58}$$

where $\times$ denotes the 3-dimesional cross product, $\mathbf{e}_i$ is the ith row of the 3-dimensional identity matrix, and $\mathbf{x}_k$ are the 3-dimensional Cartesian coordinates for atom k translated so that $\sum_k x_{k,i} = 0$ for each i. Constructed in this fashion, $\mathbf{P}_{TR}$ is a projection matrix onto the space of infinitesimal translation/rotations for the geometry given by $\mathbf{x}$, and by construction $\mathbf{P}_{TR}\mathbf{g} = 0$ when $\mathbf{P}_{TR}$ and $\mathbf{g}$ are computed at the same geometry. To eliminate the translation and rotation from a geometry optimization step, the Hessian and tangent may be modified in the following way prior to computing the step

according to eqns 5.52-5.54

$$\mathbf{H}_{prj} = (\mathbf{I} - \mathbf{P}_{TR})\,\mathbf{H}_0\,(\mathbf{I} - \mathbf{P}_{TR}) + \mathbf{P}_{TR} \tag{5.59}$$

$$\boldsymbol{\tau}_{prj} = (\mathbf{I} - \mathbf{P}_{TR})\,\boldsymbol{\tau} \tag{5.60}$$

This will ensure that $\mathbf{x}_{TS}$ has no initial rotation relative to $\mathbf{x}(t_{TS})$ prior to the microterations. At every step of the microiterations, though, $\mathbf{x}_{TS}$ will need to be rotated to remove the overall rotation relative to the current value of $\mathbf{x}(t_{ts})$, and this can be done in the same fashion used to minimize the overall rotation of the product relative to the reactant.

5.3.4 Multiple Extrema

Up until this point, the discussion has assumed that the SDRP has exactly one transition state. If there are one or more intermediate minima with a corresponding number of additional transition states, it is a trivial matter to update the appropriate equations involving the error by using the more general form of the E_{pVRE} in equation 5.2 wherever appropriate. Additionally, for the FVRC method, each additional minima/TS pair adds another $2m$ constraints and 2 optimizable values of t. As long as there are sufficient LEC per coordinate (at least 2 or 3 per coupled structure appears to be sufficient), any number of additional geometry optimizations may be coupled to the VRE minimization. However, some care must be taken to distinguish between actual transition states, and maxima that are a result of the curvature of the path rather than the curvature of the PES. The later sort of maxima will occur when the path climbs the wall of the PES. Optimization steps computed at these false maxima and their associated minima may step towards the same stationary points as one of the actual maxima/minima optimizations. This can introduce a great deal of numerical

instability into the microiterations or even result in an inconsistent FVRC Lagrangian, and should be avoided.

The false maxima may be distinguished from the actual maxima by examining the projection of the PES Hessian onto the tangent at each maxima, $\boldsymbol{\tau}^T\mathbf{H}\boldsymbol{\tau}$. If this quantity is less than zero, then the potential has the correct curvature at that point, and the optimization of that structure should be included in the FVRC optimization. If the quantity is greater than zero, then the potential does not have the correct curvature, and the corresponding maxima must be false. When this is the case, no optimization step from the false maxima and the adjacent minima that's closest in energy should be included in the optimization, and the minimization of the VRE should be sufficient to eliminate the maxima/minima in subsequent optimization steps.

5.3.5 FVRC algorithm

1. Input initial path
2. Compute VRE, VRE derivatives, ϵ, and $\boldsymbol{\sigma}$.
3. Locate and verify the t_e corresponding to the extrema (maxima and minima) along the current path, and compute the $\mathbf{x}_e$ according to the methods in section 5.3.2.
4. Set λ_σ to the RFO eigenvalue using Hessian/gradient scaled by $\boldsymbol{\sigma}^{-1/2}$ as in eq 5.20, set $\boldsymbol{\Delta C}$, $\boldsymbol{\lambda}_\kappa$, $\mathbf{y}$ and μ_ϵ to $\mathbf{0}$
5. Begin microiterations
   (a) Compute the constraints $\kappa\left(\mathbf{C}+\boldsymbol{\Delta C}\right)$ and their derivatives w.r.t. a change in the LEC
   (b) Rotate the $\mathbf{x}_e$ to remove the overall rotation relative to the current value of $\mathbf{x}\left(t_e\right)$, and compute $\boldsymbol{\Delta x}_e$
   (c) Compute $\epsilon\left(\mathbf{C}+\boldsymbol{\Delta C}\right)$ and turn on optimization of μ_ϵ if $\epsilon<0$ and $|\boldsymbol{\Delta x}|\approx 0$
   (d) Compute derivatives of the FVRC Lagrangian
   (e) Update $\boldsymbol{\Delta C}$, $\boldsymbol{\lambda}_\kappa$, $\mathbf{y}$, t_e and μ_ϵ

(f) Check augmented gradient and augmented displacement for convergence of microiterations, end microiterations if converged

(g) goto 5a

6. Update LEC for path, recompute VRE, VRE derivatives and ϵ.

7. Locate and verify the t_e corresponding to the extrema along the current path, and compute the $\mathbf{x}_e$

8. Check the gradient at the $\mathbf{x}_e$ for convergence, and end macroiterations if converged.

9. goto 4

5.4 Results and Discussion

The methods described above were implemented in Mathmatica[76]. To illustrate the behavior of the VRC methods, two analytical potential energy surfaces will be used. The first is the analytical surface of Müller and Brown[84], multiplied by a scaling a factor of 1/627.52 so that the all of the unitless energies and displacements discussed below will more closely resemble atomic units than kcal/mol. This is a 2D surface that is frequently used in the development of new methods as it is deceptively simple, containing features such as combinations of weak and stiff vibrational modes and transition states with very small basins of attraction, which are difficult for reaction path following/optimization methods to account for. The results shown all use a basis set with 9 optimizable LEC per coordinate, and the third derivatives of the PES necessary for evaluating the VRE Hessian (see eqn 5.13) were computed analytically. The line integrals for the VRE and it's derivatives were evaluated by an adaptive integrator which computes the integral on a grid, and subdivides any grid with an unacceptable error estimate. The integrals were considered converged when the absolute maximum error for each interval was less than 10^{-8}, or the absolute relative error was less than 10^{-6} the value of the integral. Since the focused VRC method requires the integration

over regions of the PES where the gradient is very close to zero, an additional termination criteria was added: the maximum number of allowed subdivisions was set to 20, if this number is exceeded, the unconverged interval is ignored (it's contribution to the quadrature is set to 0). This stopping criteria is never met in the evaluation of the integrals during the unconstrained or constrained optimizations.

The stopping criteria for the unconstrained VRC and constrained VRC algorithms were a computed RMS change to the LEC of less than 10^{-6}. The UVRC method converged in 54 iterations, with a final VRE only $1.2x10^{-6}$ higher than the VRE of the IRC. The CVRC required only 19 iterations to converge, but the final VRE was much higher at $6.0x10^{-3}$ over the IRC VRE. The FVRC method uses the convergence of the intermediates and transition states as a stopping criteria, and is considered converged when the RMS of the gradient at all intermediates and transition states is less than 10^{-6}. The FVRC method converges even more quickly than the CVRC method, requiring only 5 iterations, with a final VRE of $1.1x10^{-2}$ over the IRC VRE, which is an error of approximately 4%. To help put these VRE errors in context, the converged pathways are plotted alongside the IRC pathway in figure 5.3. The UVRC pathway is nearly indistinguishable from the IRC as the full flexibility of the basis set is used to approximate the IRC, but both the CVRC and FVRC pathways still closely follow the IRC while cutting corners in a few places.

Figure 5.4 plots the convergence of the VRE gradient (top), PES gradient at the intermediates and transition states (middle) and LEC displacement (bottom). The constraints in the CVRC and FVRC methods keep the VRE gradient from decreasing significantly even at convergence, as expected, but also demonstrates the difficulty of the UVRC optimization. By approximately iteration 25, the UVRC pathway is already lower in energy than either of the converged CVRC or FVRC, but it takes another 25 iterations for the path to find the correct parameterization to converge. The final few

iterations demonstrate quadratic convergence behavior, as one would expect with a Newton-like method. The PES gradient plot demonstrates the efficiency and appeal of the FVRC method, as the intermediates and transition states are quickly determined along with an approximation to the SDRP. Figure 5.5 shows that path at each iteration of the FVRC method, demonstrating that the transition state and intermediate near the initial path are quickly converged.

To demonstrate the performance of the FVRC method on a higher dimensional surface that's more representative of a chemical reaction, the Lennard-Jones[85] 10 atom cluster is used. In order to challenge the VRC method, Lennard-Jones network data on the Cambridge cluster database[86] was used to locate two minimum structures which were separated by two or more intermediates, with the 35th and 46th lowest energy structures satisfying that requirement. In anticipation of more intermediates and transition states, a basis set with 15 LEC per coordinate was used, and the term in equation 5.13 corresponding to the third derivative of the PES was neglected. The FVRC manages to converge to a pathway containing 4 intermediates and 5 transition states in 13 iterations, with a final VRE roughly 25% higher than the IRC VRE. Figure 5.6 shows the convergence of the VRE and PES gradients, and Figure 5.7 the energy along the pathway for selected iterations. By the 7th iteration, the energy profile along the path closely resembles the final energy.

5.5 Conclusions and future directions

The Variational Reaction Coordinate method provides a novel approach to the optimization of reaction pathways. By representing the pathway using a linear expansion in a continuous basis set, the line integral of the gradient norm and it's derivatives with respect to a change in the linear expansion coefficients provides the groundwork for constructing an iterative and variational algorithm for improving the approximation to

a steepest descent reaction pathway. Additionally, constraints to fix the relationship between the basis function parameter t and the arc length traveled by the path, as well as constraints to couple the minimization of the variational reaction energy to the minimization of one or more points along the path (corresponding to intermediates and transition states), are described. Algorithms employing these constraints are able to rapidly determine the fully-converged structure of intermediates and transition states, as well as provide a good approximation to the reaction path.

The methods described in this chapter achieve this rapid convergence at the expensive of a high per-iteration computational cost due to the necessity of using numerical methods to evaluate integrals that depend on the PES and it's derivatives. In order for this method to enjoy routine use in the study of reactions using accurate Hartree Fock and Density Functional Theory energies, alternative means to evaluate the necessary integrals must be developed to reduce the per-iteration cost to something comparable to existing ad-hoc path optimization methods.

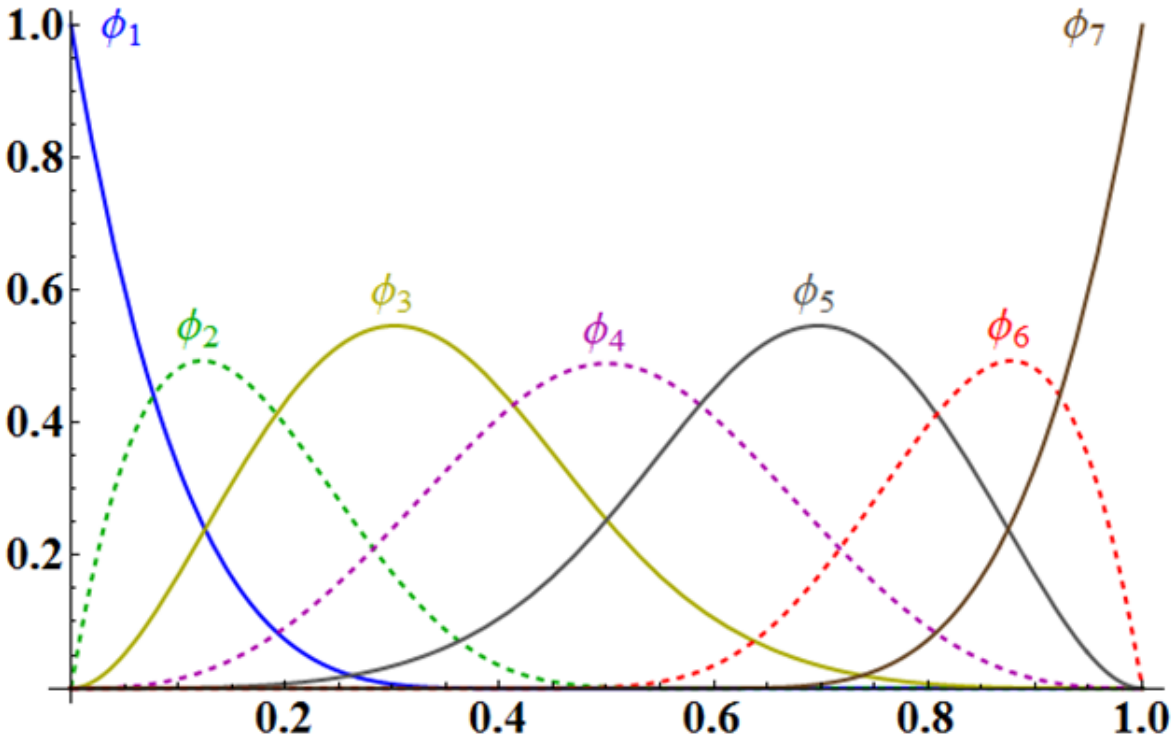


Figure 5.1: B-Spline basis with 5 internal functions ($\phi_2 - \phi_7$) and 2 capping functions (ϕ_1 and ϕ_7)

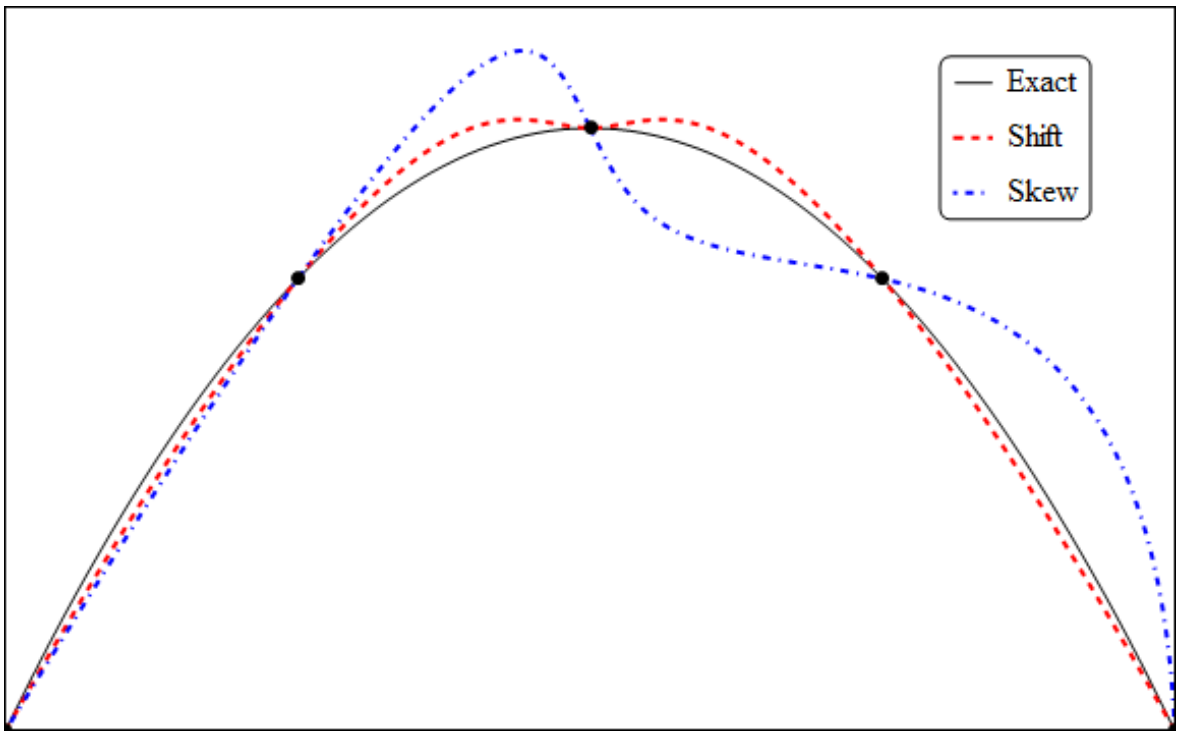


Figure 5.2: Three cubic spline curves fit to the same 5 points, using different values of t. Exact t = (0,0.25,0.5,0.7,1), Shift t = (0,0.15,0.5,0.85,1), Skew t = (0,0.1,0.4,0.7,1)

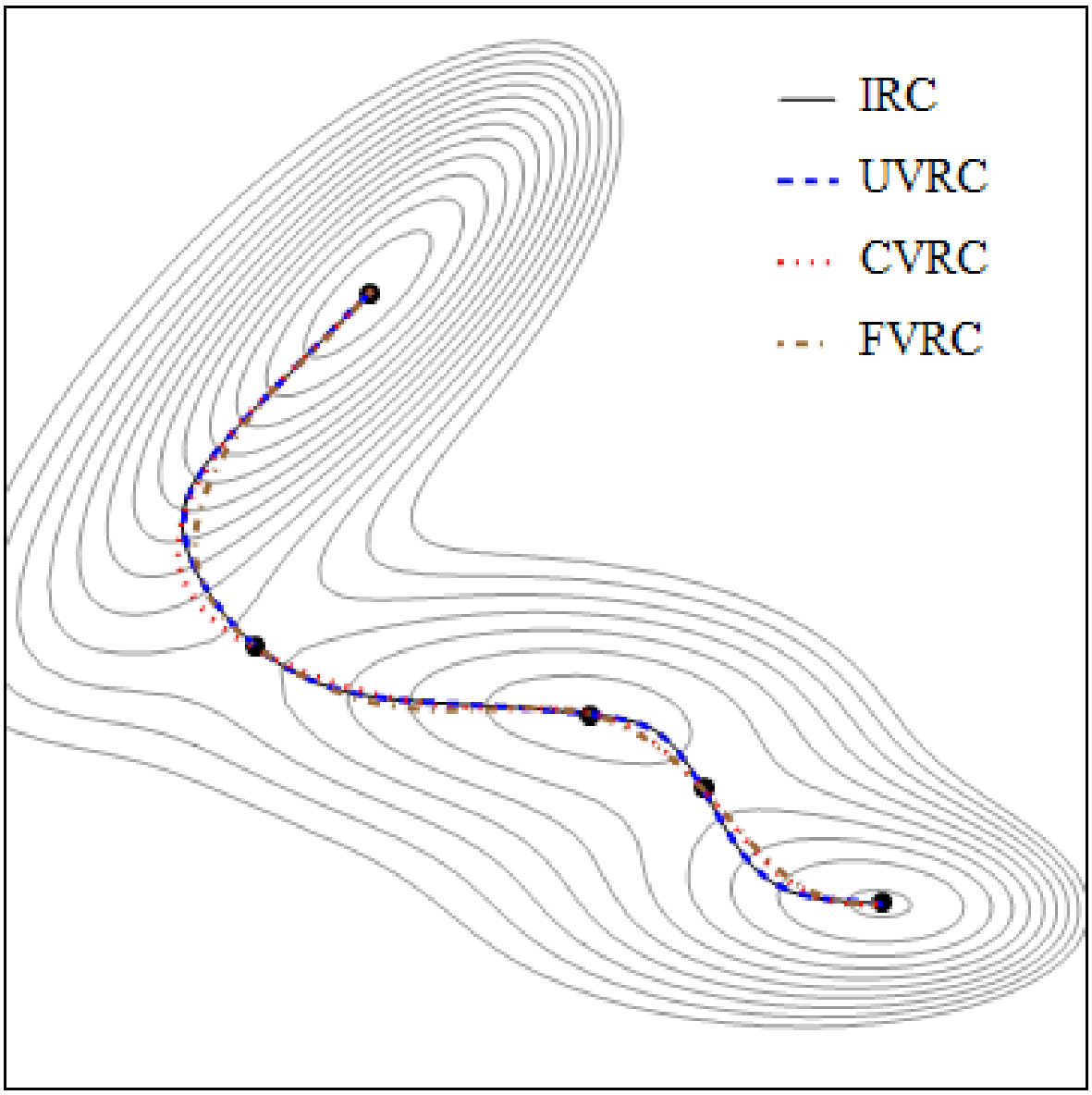


Figure 5.3: Comparison of the shape of the converged unconstrained VRC (UVRC), constrained VRC (CVRC) and focused VRC (FVRC) pathways on the Müller-Brown analytical surface. The solid curve is the IRC, and the large dots are the minima and transition states along the IRC.

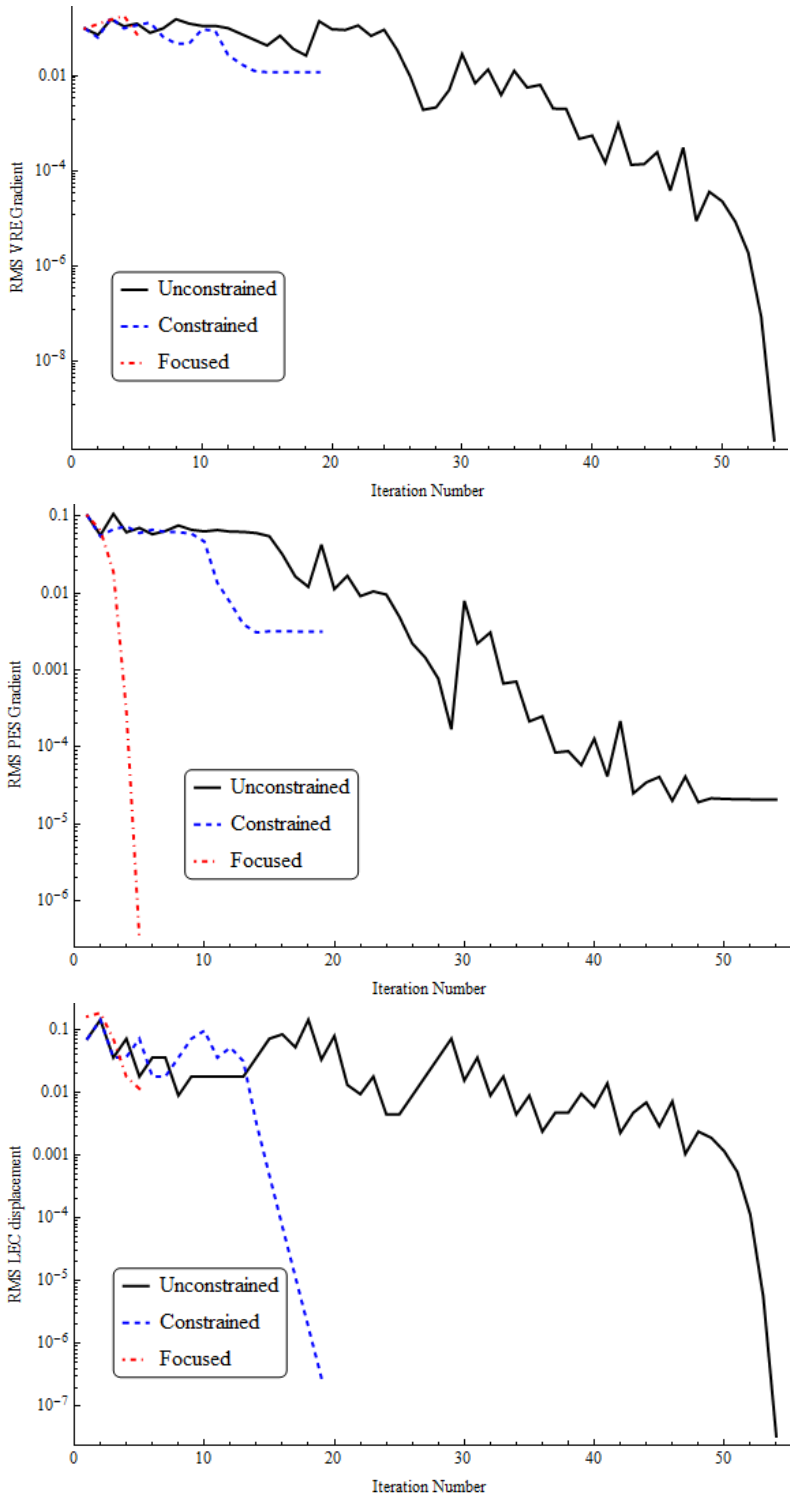


Figure 5.4: Convergence log plots for the various VRC methods on the Müller-Brown surface. **Top** RMS VRE gradient **Center** RMS PES gradient (measured at the minima and maxima along the path only) **Bottom** RMS LEC displacement

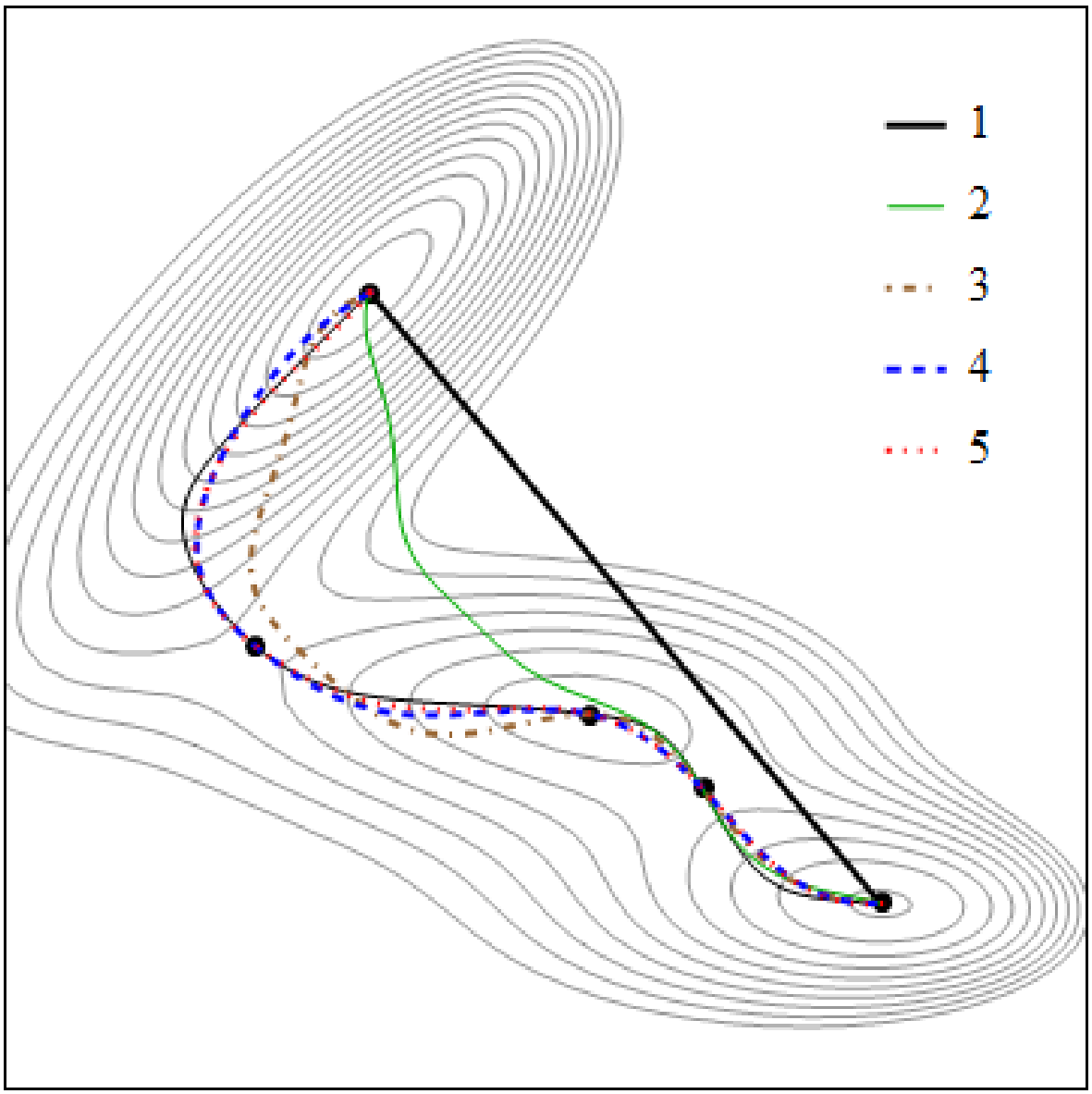


Figure 5.5: The paths at every iteration of the FVRC method on the Müller-Brown surface, compared with the IRC.

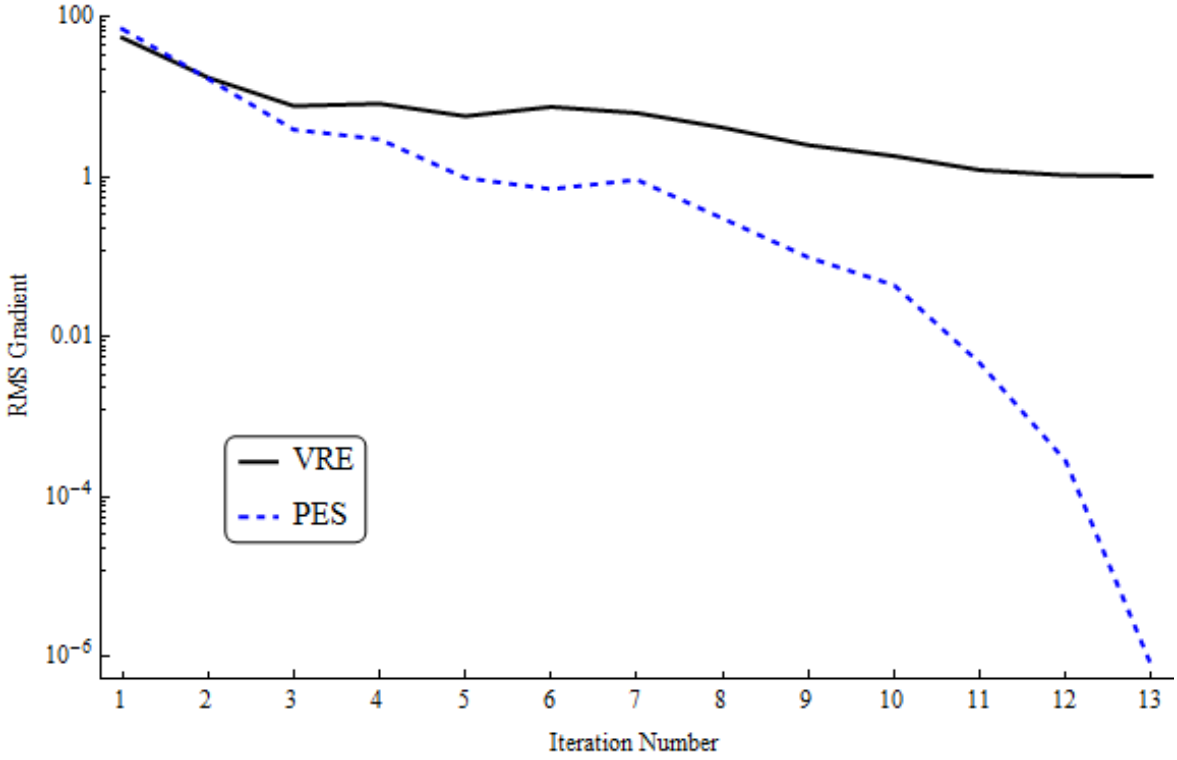


Figure 5.6: Convergence of the VRE and PES gradient using the FVRC method on the LJ10 surface.

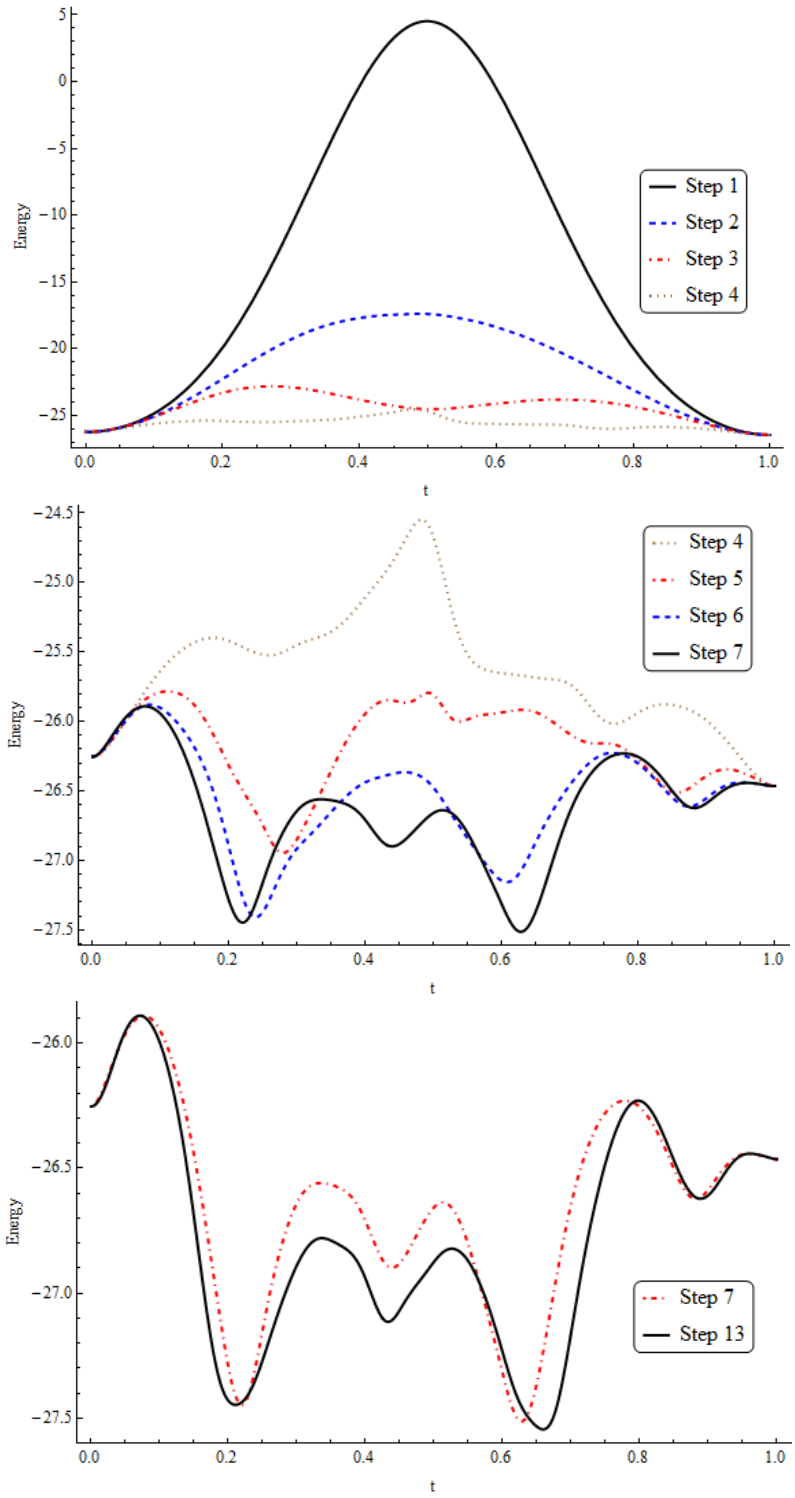


Figure 5.7: Energy plots for selected iterations of the LJ10 FVRC optimization. **Top** Iterations 1-4 **Center** Iterations 4-7, zoomed in to show detail **Bottom** Comparison of the 7th and final (13th) steps, demonstrating that by the 7th iteration, the path is already in the same region of the PES as the converged path

CHAPTER 6

INCORPORATING REDUNDANT INTERNAL COORDINATES INTO THE VARIATIONAL REACTION COORDINATE METHOD

6.1 Introduction

In the previous chapter, a variational approach to approximating the steepest descent reaction path (SDRP) was introduced as an alternative to the ad hoc "chain of states" methods [4, 77, 78, 79, 80, 5]. The chain of states methods minimize the energies of a series of discrete structures along the path, subject to fictitious forces, constraints or reparameterization schemes that ensure the distribution of structures remains uniform. These path optimization methods are primarily used to avoid the difficult problem of determining the transition state geometry, which is the starting point for more economical and accurate reaction path following methods[19, 20, 31] that determine the SDRP by walking downhill on the potential energy surface. Regardless of how it is approximated, the SDRP is defined such that every geometry along the path is a local minimum in directions orthogonal to the path. Therefore, the SDRP can be viewed as a simple approximation to the path a reaction follows as it proceeds from reactant to product.

The Variational Reaction Coordinate (VRC) method is an alternative to to the minimization of a chain of states class of path optimization methods. In the VRC method, the Variational Reaction Energy (VRE) is functional which is minimized by the SDRP[34, 33], and is defined as the line integral of the potential energy gradient

norm

$$\begin{aligned} E_{VRE} &= \int_{t_R}^{t_P} \sqrt{\frac{\partial V\left(\mathbf{x}\left(t\right)\right)}{\partial x}^{\mathrm{T}} \frac{\partial V\left(\mathbf{x}\left(t\right)\right)}{\partial x}} \sqrt{\frac{d\mathbf{x}\left(t\right)}{dt}^{\mathrm{T}} \frac{d\mathbf{x}\left(t\right)}{dt}} dt \\ &= \int_{t_R}^{t_P} \left|\mathbf{g}\left(\mathbf{x}\left(t\right)\right)\right| \left|\boldsymbol{\tau}\left(t\right)\right| dt \end{aligned} \tag{6.1}$$

where V is the potential energy, $\mathbf{x}\left(t\right)$ are the coordinates of the reaction path parameterized by t, $t_R = 0$ and $t_P = 1$ are the parameter values corresponding to the reactant and product structures, respectively. For clarity, $\mathbf{g}$ and $\boldsymbol{\tau}$ are used as shorthand for the gradient of the potential, and the tangent to the path. The VRC method represents the path as a linear expansion of basis functions:

$$\begin{aligned} x_i\left(t\right) = \sum_{\mu} C_{i\mu}\phi_{\mu}\left(t\right) \quad &\rightarrow \quad \frac{\partial x_i\left(t\right)}{\partial C_{i\mu}} = \phi_{\mu}\left(t\right) \\ \tau_i\left(t\right) = \sum_{\mu} C_{i\mu}\frac{d\phi_{\mu}\left(t\right)}{dt} \quad &\rightarrow \quad \frac{\partial \tau_i\left(t\right)}{\partial C_{i\mu}} = \frac{d\phi_{\mu}\left(t\right)}{dt} \end{aligned} \tag{6.2}$$

where the Roman indices are over the $n_{crt} = 3 \times n_{atoms}$ Cartesian coordinates, the Greek indicies are over the basis functions, and the ϕ are quartic B-Spline functions. For a given choice of basis, the linear expansion coefficients (LEC, $C_{i\mu}$) define the path, and derivatives of the VRE with respect to a change in the LEC may be defined as

$$\gamma_{i\mu} = \int_{t_R}^{t_P} \left(\frac{|\boldsymbol{\tau}|}{|\mathbf{g}|} \sum_a H_{ia} g_a \phi_{\mu} + \frac{|\mathbf{g}|}{|\boldsymbol{\tau}|} \tau_i \frac{d\phi_{\mu}}{dt} \right) dt \tag{6.3}$$

$$\eta_{i\mu j\nu} = \int_{t_R}^{t_P} \left(\begin{array}{l} \frac{|\boldsymbol{\tau}|}{|\mathbf{g}|} \left(\left(Tg + HH\right)_{ij} - \frac{\sum_{a,b} H_{ia} g_a H_{jb} g_b}{|\mathbf{g}|^2} \right) \phi_{\mu}\phi_{\nu} + \\ \frac{\sum_a H_{ia} g_a \tau_i}{|\mathbf{g}||\boldsymbol{\tau}|} \phi_{\mu} \frac{d\phi_{\nu}}{dt} + \frac{\tau_i \sum_a H_{ja} g_a}{|\mathbf{g}||\boldsymbol{\tau}|} \frac{d\phi_{\mu}}{dt} \phi_{\nu} + \\ |\mathbf{g}| \left(\frac{\delta_{ij}}{|\boldsymbol{\tau}|} - \frac{\tau_i \tau_j}{|\boldsymbol{\tau}|^3} \right) \frac{d\phi_{\mu}}{dt} \frac{d\phi_{\nu}}{dt} \end{array} \right) dt \tag{6.4}$$

$$Q_{VRE}\left(\mathbf{C} + \ \mathbf{C}\right) = E\left(\mathbf{C}\right) + \Delta\mathbf{C}^{\mathrm{T}}\boldsymbol{\gamma}\left(\mathbf{C}\right) + \frac{1}{2}\Delta\mathbf{C}^{\mathrm{T}}\boldsymbol{\eta}\left(\mathbf{C}\right)\Delta\mathbf{C} \tag{6.5}$$

where **H** is the potential energy Hessian, and **Tg** is the product of the third derivative of the potential energy with the potential energy gradient. Additionally, a positive definite shift matrix $\boldsymbol{\sigma}_{i\mu j\nu}$ is used to shift $\boldsymbol{\eta}$ so that it is positive definite while attempting to account for the curvilinear relationship between the LEC and the gradient, and a set of n constraints κ_α to define the relationship between the parameter t and the arc length of the path:

$$\sigma_{i\mu j\nu} = \int_{t_R}^{t_P} \delta_{ij} \frac{d\phi_\mu}{dt} \frac{d\phi_\nu}{dt} dt \tag{6.6}$$

$$\kappa_\alpha = S(t_{\alpha-1}, t_\alpha) - S(t_\alpha, t_{\alpha+1}) \qquad 1 \leq \alpha \leq n \tag{6.7}$$

where S is the arc length between two points along the path, and the t_α are evenly spaced to give $n_{bas} + 1$ segments along the path, where n_{bas} is the number of basis functions used to expand each coordinate

$$S(t_1, t_2) = \int_{t_1}^{t_2} |\boldsymbol{\tau}(t)| \, dt \tag{6.8}$$

$$t_\alpha = \frac{\alpha}{n+1} \qquad 0 \leq \alpha \leq n+1 \tag{6.9}$$

During every iteration of the constrained VRC method, E, $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ and $\boldsymbol{\sigma}$ are computed via numerical integration, and the following Lagrangian is minimized

$$\mathcal{L}_{CVRC} = Q_{VRE}(\mathbf{C} + \boldsymbol{\Delta}\mathbf{C}) - \frac{1}{2}\lambda_\sigma \Delta\mathbf{C}^{\mathrm{T}} \boldsymbol{\sigma}(\mathbf{C}) \Delta\mathbf{C} + \sum_\alpha \lambda_\alpha \kappa_\alpha (\mathbf{C} + \boldsymbol{\Delta}\mathbf{C}) \tag{6.10}$$

with respect to the change in the LEC ($\boldsymbol{\Delta}$C) and the Lagrange multipliers for the constraints (λ_α), while λ_σ is determined by the RFO method[54]. This is done using microterations so that the curvilinear relationship of the constraints with respect to the change in the LEC can be accounted for and every macro-iteration begins with a path that satisfies $\kappa_\alpha = 0$ for all α. This constrained approach (CVRC) converges

more rapidly than the unconstrained approach, and if coupled with optimization of the points along the path that are stationary w.r.t. t in order to focus the optimization in a direction most likely to contain the SDRP (FVRC), both a good approximation to the SDRP and the converged geometries of any transition states or intermediate structures along the path can be determined simultaneously with only a small number of iterations required for convergence. This algorithmic efficiency comes at the expense of a high per-iteration cost, due to the necessity of evaluating the VRE and it's derivatives by numerical quadrature methods, and previous work was limited to demonstrations on analytical test surfaces. This chapter will focus on the development of additional methods to improve the applicability of the FVRC method to the study of chemical reactions, as well as to reduce the per-iteration cost to something comparable to existing ad-hoc path optimization approaches.

The VRC methods described above were derived in chapter **??** for the case when both path and the derivatives of the potential energy are all expressed in Cartesian coordinates. For chemical potential energy surfaces, internal coordinates consisting of a redundant combination of bond stretches, valence angle bends, and dihedral angle torsions typically provide a much more natural description of the relative motion of atoms[11], which results in less coupling between coordinates, and a potential energy surface that is likely to be more constant in the degrees of freedom not directly involved with a reaction. Incorporating these redundant internal coordinates (RIC) into the VRC method should not only have a positive effect on the efficiency of the algorithm, but should also improve the accuracy of methods to approximate the PES (i.e. through Hessian updating) which could help reduce the per-iteration cost of the VRC method in the future.

Since the potential surface and the relationship between Cartesians and the RIC are both defined with respect to the Cartesian coordinates, and since there is not necessarily

a one to one relationship between an RIC geometry and a Cartesian geometry, the present work is limited to methods which can be implemented using a path in Cartesian coordinates. This avoids the introduction of any possible error/uncertainty in the path due to back-transformation procedures, but may also limit the overall benefit of using RIC as some of the more curvilinear motions from the bends and torsions cannot be fully utilized. Three approaches to incorporate redundant internal coordinates into the FVRC method are introduced, and preliminary data are presented and discussed to motivate further investigation:

- Express the FVRC coupling constraints in terms of redundant internal coordinate differences in order to avoid having to deal with a separate rotational alignment step
- Applying the methods developed for the CVRC method to optimize a least RIC length pathway by minimizing the arc length expressed in RIC. The resulting path should be a better initial guess for the VRC method than a linear Cartesian pathway
- Define the VRE and it's derivatives in terms of an interpolated RIC PES.

6.2 RIC Coupling Constraints for the FVRC Method

In the Focused Variational Reaction Coordinate method, coupling constraints of the form

$$0 = \Delta_i x_{ext} = x_i(t_{ext}) - x_{ext,i} \qquad 1 \leq i \leq n_{crt} \tag{6.11}$$

are added to the CVRC Lagrangian as the additional term $\mathcal{F} = \left(\frac{1}{2}\Delta\mathbf{x} + \mathbf{y}\right)^T \Delta\mathbf{x}$, where the $\mathbf{y}$ are Lagrange multipliers, in order to require that the pathway passes through the one or more structures $\mathbf{x}_{ext}$. These $\mathbf{x}_{ext}$ determined prior to the microiterations by

computing a single geometry optimization step at a point along the path corresponding to an energy maximum (TS) or minimum (intermediate). In order to improve the stability of the optimization, $\mathbf{x}_{ext}$ has to be rotated during each microiteration in order to minimize the overall rotation between $\mathbf{x}_{ext}$ and the current value of $\mathbf{x}\left(t_{ext}\right)$. It would be advantageous to use $\mathcal{F}_q = \left(\frac{1}{2}\Delta\mathbf{q} + \mathbf{y}_q\right)^T \Delta\mathbf{q}$, where $\mathbf{q}$ are internal coordinates, but if the internal coordinates have any redundancies in them the x,y second derivative of F_q will be overdetermined, resulting in a singular augmented Hessian. The solution is to define the derivatives of the coupling constraints only in the locally non-redundant space defined as the first $n_{act} = n_{crt} - 6$ left-singular vectors of the B-Matrix, $\mathbf{U}_{act}$. The RIC coupling constraints take the form

$$0 = \Delta_a q_{ext} = \sum_b \left[\mathbf{U}_{act}\right]_{ab} \left(q_b(t_{ext}) - q_{ext,b}\right) \qquad 1 \leq a \leq n_{act} \tag{6.12}$$

where 2π is added or subtracted to the differences corresponding to dihedral coordinates as necessary to account for the discontinuity at +/- π. To compute the derivatives of 6.12, the reduced Lagrange multipliers, B-matrix, B-matrix derivative and RIC displacements are defined as follows

$$\Delta\mathbf{q}_r = \mathbf{U}_{act}^T \Delta\mathbf{q}_{ext} \qquad \mathbf{y}_r = \mathbf{U}_{act}^T \mathbf{y}_q \tag{6.13}$$

$$\mathbf{B}_r = \mathbf{U}_{act}^T \mathbf{B} \qquad \frac{\partial \mathbf{B}_r}{\partial \mathbf{x}} = \mathbf{U}_{act}^T \frac{\partial \mathbf{B}}{\partial \mathbf{x}} \tag{6.14}$$

where the r subscript corresponds to the reduced quantity. This makes the incorrect assumption that $\mathbf{U}_{act}$ is constant with respect to a change in the Cartesian coordinates, but since the x-derivative of $\mathbf{U}_{act}$ term will vanish in $\mathbf{B}_r$ when $\Delta\mathbf{q}$ is zero, it should be safe to neglect. The $\mathbf{y}_r$, t_{ext} and LEC derivatives of $\mathcal{F}_q$ are

$$\frac{\partial \mathcal{F}_q}{\partial y_a} = \Delta_a q_{ext} \tag{6.15}$$

$$\frac{d\mathcal{F}_q}{dt_{ext}} = \left(\Delta \mathbf{q}_r + \mathbf{y}_r\right)^T \mathbf{B}_r \boldsymbol{\tau}_{ext} \tag{6.16}$$

$$\frac{\partial \mathcal{F}_q}{\partial C_{i\mu}} = \sum_a \left(\Delta \mathbf{q}_r + \mathbf{y}_r\right)_a \left[\mathbf{B}_r\right]_{ai} \phi_\mu \left(t_{ext}\right) \tag{6.17}$$

$$\frac{d\partial \mathcal{F}_q}{dt_{ext}\partial y_a} = \sum_i \left[\mathbf{B}_r\right]_{ai} \left[\boldsymbol{\tau}_{ext}\right]_i \tag{6.18}$$

$$\left[\mathbf{M}_{xx}\right]_{ij} = \frac{\partial^2 \mathcal{F}_q}{\partial x_i \partial x_j} = \sum_b \left[\mathbf{B}_r\right]_{bi} \left[\mathbf{B}_r\right]_{bj} + \left(\Delta \mathbf{q}_r + \mathbf{y}_r\right)_b \frac{\partial \left[\mathbf{B}_r\right]_{bi}}{\partial x_j} \tag{6.19}$$

$$\frac{d^2 \mathcal{F}_q}{dt_{ext}^2} = \boldsymbol{\tau}_{ext}^T \mathbf{M}_{xx} \boldsymbol{\tau}_{ext} + \left(\Delta \mathbf{q}_r + \mathbf{y}_r\right)^T \mathbf{B}_r \frac{d\boldsymbol{\tau}_{ext}}{dt_{ext}} \tag{6.20}$$

$$\frac{d\partial \mathcal{F}_q}{dt_{ext}\partial C_{i\mu}} = \left[\mathbf{M}_{xx}\boldsymbol{\tau}_{ext}\right]_i \phi_\mu \left(t_{ext}\right) + \left[\left(\Delta \mathbf{q}_r + \mathbf{y}_r\right)^T \mathbf{B}_r\right]_i \frac{d\phi_\mu \left(t_{ext}\right)}{dt} \tag{6.21}$$

$$\frac{\partial^2 \mathcal{F}_q}{\partial C_{i\mu}\partial y_a} = \left[\mathbf{B}_{ai}\right] \phi_\mu \left(t_{ext}\right) \tag{6.22}$$

$$\frac{\partial^2 \mathcal{F}}{\partial C_{i\mu}\partial C_{j\nu}} = \left[\mathbf{M}_{xx}\right] \phi_\mu \left(t_{ext}\right) \phi_\nu \left(t_{ext}\right) \tag{6.23}$$

With the derivatives defined in this fashion, $\mathcal{F}_q$ can replace $\mathcal{F}$ in the FVRC Lagrangian.

6.3 Least Length Path

6.3.1 RIC Arc Length

The arc length formula (eqn 6.8) generalizes the concept of length to things that are curved in nature, like the length traveled by a reaction path in Cartesian coordinates. It can also be used to compare lengths of curves in curvilinear spaces, which are spaces that have a variable metric. A metric defines the infinitesimal relationship between a change in the coordinates, and the change in the length traveled. A simple conceptual example of this is computing distances on the surface of a sphere. In this case, the

curvilinear coordinate system could be a pair of angles (i.e. a spherical coordinate system with a fixed radius), and the metric would define how to compute the arc of a circle on the cross section of the sphere containing two points.

For molecular geometries, the RIC provide a curvilinear coordinate set that is a more naturally uncoupled representation of the relative motions of atoms, and Wilson's B-Matrix ($B_{ai} = \frac{\partial q_a}{\partial x_i}$, with q the RIC, and x the Cartesian coordinates) defines the metric. To determine the RIC length of a path defined in Cartesian coordinates, the following arc length formula can be used instead

$$S_{RIC}(t_1, t_2) \quad = \quad \int_{t_1}^{t_2} \sqrt{\boldsymbol{\tau}^T \mathbf{G}_C \boldsymbol{\tau}} dt \tag{6.24}$$

where $\mathbf{G}_C = \mathbf{B}^T\mathbf{B}$, $\mathbf{B}$ depends on $\mathbf{x}(t)$, and $\boldsymbol{\tau}$ depends on t. $\mathbf{G}_C$ can be thought of as the n_{crt} x n_{crt} inner product version of Wilson's G-matrix $\mathbf{G}_Q = \mathbf{B}\mathbf{B}^T$, which is commonly used in redundant internal coordinate transformations. Any path between reactant and product that minimizes S_{RIC} will have the smallest possible overall change in the redundant internals. This should provide a more reasonable initial pathway for optimization than the linear path in Cartesians by eliminating nonphysical or unrealistic configurations, such as those with atomic collisions due to very small bonds or angles. By modifying the CVRC method developed in the previous chapter, an arc length formula like S_{RIC} can be minimized in order to determine a path in Cartesian coordinates that is ready to be optimized by VRC or some other path optimization method.

Since $\mathbf{G}_C$ is defined as a matrix in Cartesian coordinates, no additional consideration of the possible redundancy in the RIC coordinate system is necessary beyond the requirement that the selection of internal coordinates is capable of fully describing all of the internal motions of the molecule (i.e. the B-Matrix has n_{act} linearly inde-

pendent rows). However, even when the RIC are well chosen the matrix $\mathbf{G}_C$ will be singular due to the external degrees of freedom, and therefore derivatives of 6.24 with respect to a change in the LEC will also be singular. S_{RIC} will have to be modified in order to ensure numerical stability of a Newton optimization. The most direct route would be to add the projector $\mathbf{P}_C^{\perp} = \mathbf{I} - \mathbf{B}^{-}\mathbf{B}$ to $\mathbf{G}_C$, where $\mathbf{I}$ is the identity matrix, and $\mathbf{B}^{-}$ indicates the pseudoinverse of $\mathbf{B}$, but this would add significant cost to the evaluation of eq 6.24, and would complicate the derivation of the derivatives necessary to implement a VRC-like method for minimizing S_{RIC}. Since $\mathbf{P}_C^{\perp}$ spans the external degrees of freedom, overall translation and rotation, the projector may be conveniently constructed without inversion of the B-matrix:

$$\mathbf{P}_{TR} = \sum_{i}^{3} \left(\frac{1}{n_{atoms}} \mathbf{t}_i \mathbf{t}_i^T + \mathbf{r}_i \left(\sum_{j}^{3} \mathbf{r}_j^T \mathbf{r}_j \right)^{-1} \mathbf{r}_i^T \right) = \frac{\mathbf{T}\mathbf{T}^T}{n_{atoms}} + \mathbf{R} \left(\mathbf{R}^T \mathbf{R} \right)^{-1} \mathbf{R}^T \quad (6.25)$$

where the portion of these vectors corresponding to the kth atom are given by

$$\mathbf{t}_{i,k} = \mathbf{e}_i \quad (6.26)$$

$$\mathbf{r}_{i,k} = \mathbf{x}_k \times \mathbf{e}_i \quad (6.27)$$

where $\times$ denotes the 3-dimesional cross product, $\mathbf{e}_i$ is the ith row of the 3-dimensional identity matrix, and $\mathbf{x}_k$ are the 3-dimensional Cartesian coordinates for atom k translated so that $\sum_k x_{k,i} = 0$ for each i. It can be verified numerically that $\frac{\partial}{\partial \mathbf{x}} \left(\boldsymbol{\tau}^T \mathbf{P}_C^{\perp} \boldsymbol{\tau} \right) = \frac{\partial}{\partial \mathbf{x}} \left(\boldsymbol{\tau}^T \mathbf{P}_{TR} \boldsymbol{\tau} \right)$. Adding this projector to eqn 6.24 results in the following modified RIC arc length, which is more suitable for optimization

$$S_{RIC+} = \int_{t_R}^{t_P} \sqrt{\boldsymbol{\tau}^T \left(\mathbf{G}_C + \mathbf{P}_{TR} \right) \boldsymbol{\tau}} dt \quad (6.28)$$

The projector in 6.25 has two components, the first corresponding to overall translation, and the second corresponding to infinitesimal rotations. As long as the initial path contains no overall translation, the addition of the first term does nothing but eliminate singularities in the second derivatives of S_{RIC+}. The second term, on the other hand, penalizes the introduction of overall rotation to the path in order to further reduce the internal coordinate arc length. Without it, the second derivatives of S_{RIC+} would not necessarily be singular, but they will likely be very ill-conditioned, and the converged path may have some aesthetically unpleasant rocking motions included during visualization.

6.3.2 Derivatives and Modified RIC Length Lagrangian

The first step in applying the CVRC methods to minimizing eqn 6.28 is to compute the derivatives with respect to a change in the LEC. For notational convenience, let $\mathbf{M}_S = \mathbf{G}_C + \mathbf{P}_{TR}$, and $|\boldsymbol{\tau}_{\boldsymbol{M}}| = \sqrt{\boldsymbol{\tau}^T \mathbf{M}_S \boldsymbol{\tau}}$. The first derivative is given by

$$\frac{\partial S_{RIC+}}{\partial C_{i\mu}} = \int_{t_R}^{t_P} \left(\phi_\mu \frac{\partial}{\partial x_i} + \frac{d\phi_\mu}{dt} \frac{\partial}{\partial \tau_i} \right) |\boldsymbol{\tau}_{\boldsymbol{M}}| \, dt \tag{6.29}$$

$$\frac{\partial}{\partial x_i} |\boldsymbol{\tau}_{\boldsymbol{M}}| = \frac{\sum_{k,l} \tau_k \left(\frac{\partial [\mathbf{G}_C]_{kl}}{\partial x_i} + \frac{\partial [\mathbf{P}_{TR}]_{kl}}{\partial x_i} \right) \tau_l}{2 \, |\boldsymbol{\tau}_{\boldsymbol{M}}|} \tag{6.30}$$

$$\frac{\partial}{\partial \tau_i} |\boldsymbol{\tau}_{\boldsymbol{M}}| = \frac{\sum_k \tau_k \left([\mathbf{M}_S]_{ik} \right)}{|\boldsymbol{\tau}_{\boldsymbol{M}}|} \tag{6.31}$$

with the x-derivatives of $\mathbf{G}_C$ constructed straightforwardly as

$$\frac{\partial G_{kl}}{\partial x_i} = \sum_a B_{ak} \frac{\partial B_{al}}{\partial x_i} + \frac{\partial B_{ak}}{\partial x_i} B_{al} \tag{6.32}$$

while the x-derivatives of $\mathbf{P}_{TR}$ are a little more involved. The first term in 6.25 is a projector onto the overall translation, and so is constant. The second term is a projector onto the infinitesimal rotation, and its derivative can be computed most conveniently

as the product $\boldsymbol{\tau}^T \frac{\partial \mathbf{P}_R}{\partial \mathbf{x}} \boldsymbol{\tau}$, where $\mathbf{P}_R = \mathbf{R}\left(\mathbf{R}^T\mathbf{R}\right)^{-1}\mathbf{R}^T$. Start by defining the length 3 vector $\boldsymbol{\tau}^+$, and the length n_{crt} vector $\boldsymbol{\tau}^-$

$$\boldsymbol{\tau}^+ = \boldsymbol{\tau}^T \left(\mathbf{I} + \mathbf{P}_R\right) \mathbf{R}\left(\mathbf{R}^T\mathbf{R}\right)^{-1} \tag{6.33}$$

$$\boldsymbol{\tau}^- = \left(\mathbf{I} - \mathbf{P}_R\right)\boldsymbol{\tau} \tag{6.34}$$

then the portion of $\boldsymbol{\tau}^T \frac{\partial \mathbf{P}_R}{\partial \mathbf{x}} \boldsymbol{\tau}$ corresponding to atom k is given by

$$\left[\boldsymbol{\tau}^T \frac{\partial \mathbf{P}_R}{\partial \mathbf{x}} \boldsymbol{\tau}\right]_k = 2\boldsymbol{\tau}^+ \times \left[\boldsymbol{\tau}^-\right]_k \tag{6.35}$$

The second derivatives are derived in a similar fashion

$$\frac{\partial^2 S_{RIC+}}{\partial C_{i\mu}\partial C_{j\nu}} = \int_{t_R}^{t_P} \left(\phi_\mu \phi_\nu \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} + \frac{d\phi_\mu}{dt}\phi_\nu \frac{\partial}{\partial \tau_i}\frac{\partial}{\partial x_j} + \phi_\mu \frac{d\phi_\nu}{dt}\frac{\partial}{\partial x_i}\frac{\partial}{\partial \tau_j} + \frac{d\phi_\mu}{dt}\frac{d\phi_\nu}{dt}\frac{\partial}{\partial \tau_i}\frac{\partial}{\partial \tau_j} \right) \left|\boldsymbol{\tau_M}\right| dt \tag{6.36}$$

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\left|\boldsymbol{\tau_M}\right| = \frac{\sum_{k,l} \tau_k \left(\frac{\partial^2 [\mathbf{G}_C]_{kl}}{\partial x_i \partial x_j} + \frac{\partial^2 [\mathbf{P}_{TR}]_{kl}}{\partial x_i \partial x_i} \right) \tau_l}{2\left|\boldsymbol{\tau_M}\right|} - \frac{\partial \left|\boldsymbol{\tau_M}\right|}{\partial x_i}\frac{\partial \left|\boldsymbol{\tau_M}\right|}{\partial x_j}\frac{1}{\left|\boldsymbol{\tau}_M\right|} \tag{6.37}$$

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial \tau_j}\left|\boldsymbol{\tau_M}\right| = \frac{\sum_{k} \tau_k \left(\frac{\partial [\mathbf{G}_C]_{kj}}{\partial x_i} + \frac{\partial [\mathbf{P}_{TR}]_{kj}}{\partial x_i} \right)}{2\left|\boldsymbol{\tau_M}\right|} - \frac{\partial \left|\boldsymbol{\tau_M}\right|}{\partial x_i}\frac{\partial \left|\boldsymbol{\tau_M}\right|}{\partial \tau_j}\frac{1}{\left|\boldsymbol{\tau}_M\right|} \tag{6.38}$$

$$\frac{\partial}{\partial \tau_i}\frac{\partial}{\partial \tau_j}\left|\boldsymbol{\tau_M}\right| = \frac{[\mathbf{M}_S]_{ij}}{\left|\boldsymbol{\tau_M}\right|} - \frac{\partial \left|\boldsymbol{\tau_M}\right|}{\partial \tau_i}\frac{\partial \left|\boldsymbol{\tau_M}\right|}{\partial \tau_j}\frac{1}{\left|\boldsymbol{\tau}_M\right|} \tag{6.39}$$

$$\frac{\partial^2 G_{kl}}{\partial x_i \partial x_j} = \sum_a 2\frac{\partial B_{ak}}{\partial x_i}\frac{\partial B_{al}}{\partial x_j} + B_{ak}\frac{\partial^2 B_{al}}{\partial x_i \partial x_j} + \frac{\partial^2 B_{ak}}{\partial x_i \partial x_j}B_{al} \tag{6.40}$$

The absolute accuracy of the second derivatives is not as essential for an optimization process as the accuracy of the first derivatives, and numerical tests seem to indicate that the contribution from both the x-derivative of $\mathbf{P}_{TR}$ and the τ-derivative

of $\boldsymbol{\tau}^T \frac{\partial \mathbf{P}_R}{\partial \mathbf{x}} \boldsymbol{\tau}$ tends to be small, so these terms will be neglected for now and the analytical derivation will be left for future work. Likewise, without analytical expressions available for the second derivative of the B-matrix, the cost of the second term in eqn 6.40 is prohibitive, and this term is also neglected.

In addition to the LEC derivatives for the S_{RIC+}, the following shift matrix is used in the place of the overlap of the basis set derivatives used in the VRC method

$$[\boldsymbol{\sigma}_S]_{i\mu j\nu} = \int_{t_R}^{t_P} \left(\frac{d\phi_\mu}{dt} \frac{d\phi_\nu}{dt} \frac{[\mathbf{M}_S]_{ij}}{|\boldsymbol{\tau}_M|} \right) dt \tag{6.41}$$

The assembled Lagrangian for the minimization of the modified RIC arc length is

$$\mathcal{L}_S = Q_S \left(\mathbf{C} + \mathbf{\Delta C} \right) - \frac{1}{2} \lambda_\sigma \Delta \mathbf{C}^{\mathrm{T}} \boldsymbol{\sigma}_S \left(\mathbf{C} \right) \Delta \mathbf{C} + \sum_\alpha \lambda_\alpha \kappa_\alpha \left(\mathbf{C} + \mathbf{\Delta C} \right) \tag{6.42}$$

note that the arc length constraints are still computed using the standard Cartesian arc length as outlined in eqns 6.7-6.9, as the path is still expressed in Cartesians, and the role of the κ functionals is to define a particular parameterization of the path independent of the functional being minimized.

6.3.3 Improved $\boldsymbol{\lambda_\sigma}$ Initialization

In the CVRC method described in chapter **??**, λ_σ was determined once per macroiteration prior to beginning the microiterations by applying the RFO method to the unconstrained VRE derivatives. While this approach did result in a working algorithm, λ_σ did not display the correct convergence behavior (i.e. it did not decrease to zero at convergence) and the resulting over-correction was suspected to play a role in the slowdown observed in the method in the final few iterations. Here, an iterative method to initialize the values of both λ_σ and the $\boldsymbol{\lambda}_\kappa$ is described that will result in the correct

convergence behavior. This is accomplished by alternately computing λ_σ followed by updating the $\boldsymbol{\lambda}_\kappa$ until no further change in the $\boldsymbol{\lambda}_\kappa$ is observed.

To begin, the $\boldsymbol{\lambda}_\kappa$ are set to zero. Each iteration begins by computing the first and second derivative of the S_{RIC} Lagrangian with λ_σ set to 0

$$\frac{\partial \mathcal{L}_S}{\partial \mathbf{C}} = \frac{\partial S_{RIC+}}{\partial \mathbf{C}} + \sum_\alpha \lambda_\alpha \frac{\partial \kappa_\alpha}{\partial \mathbf{C}} \tag{6.43}$$

$$\frac{\partial^2 \mathcal{L}_S}{\partial \mathbf{C}^2} = \frac{\partial^2 S_{RIC+}}{\partial \mathbf{C}^2} + \sum_\alpha \lambda_\alpha \frac{\partial^2 \kappa_\alpha}{\partial \mathbf{C}^2} \tag{6.44}$$

and using these derivatives along with $\boldsymbol{\sigma}_S$ to compute λ_σ. Then, the following system of equations can be constructed

$$\begin{bmatrix} \mathbf{X} & \mathbf{Y}^T \\ \mathbf{Y} & \mathbf{0} \end{bmatrix} \begin{pmatrix} \Delta \mathbf{C} \\ \Delta \boldsymbol{\lambda}_\kappa \end{pmatrix} = \begin{pmatrix} \mathbf{w} \\ \mathbf{v} \end{pmatrix} \tag{6.45}$$

$$\mathbf{X} = \frac{\partial^2 \mathcal{L}_S}{\partial \mathbf{C}^2} - \lambda_\sigma \boldsymbol{\sigma}_S \qquad \mathbf{w} = \frac{\partial \mathcal{L}_S}{\partial \mathbf{C}} \tag{6.46}$$

$$\mathbf{Y} = \frac{\partial \kappa_\alpha}{\partial \mathbf{C}} \qquad \mathbf{v} = \boldsymbol{\kappa} \tag{6.47}$$

and the Schur complement of $\mathbf{X}$ can be used to solve for $\Delta \boldsymbol{\lambda}_\kappa$

$$\Delta \boldsymbol{\lambda}_\kappa = -\left(\mathbf{Y}^T \mathbf{X}^{-1} \mathbf{Y}\right)^{-1} \left(\mathbf{Y}^T \mathbf{X}^{-1} \mathbf{w} - \mathbf{v}\right) \tag{6.48}$$

which is used to update λ_α. This process repeats until the RMS of $\Delta \boldsymbol{\lambda}_\kappa$is less than 10^{-6}. This method also shows convergent behavior when additional constraints are added to $\mathbf{Y}$, $\Delta \boldsymbol{\lambda}$ and $\mathbf{v}$, so long as $\mathbf{v}$ is small. These conditions are always met for the arc length constraints, but may not be met by the coupling constraints in the FVRC method.

6.3.4 RIC Arc Length Minimization Algorithm

1. Input initial (linear Cartesian) path
2. Compute S_{RIC+}, S_{RIC+} derivatives and $\boldsymbol{\sigma}_S$
3. Initialize λ_σ and the $\boldsymbol{\lambda}_\kappa$ according to sec 6.3.3
4. Begin microiterations
   (a) Compute the constraints $\kappa\left(\mathbf{C}+\boldsymbol{\Delta}\mathbf{C}\right)$ and their derivatives w.r.t. a change in the LEC
   (b) Compute the (augmented) $\boldsymbol{\Delta}\mathbf{C}$ and $\boldsymbol{\lambda}_\kappa$ derivatives of the $\mathcal{L}_S$
   (c) Update $\boldsymbol{\Delta}\mathbf{C}$ and $\boldsymbol{\lambda}_\kappa$ using Newton's method to produce an augmented displacement
   (d) Check augmented gradient and augmented displacement for convergence of microiterations, end microiterations if converged
   (e) goto 4a
5. Update LEC for path, and recompute S_{RIC+}, S_{RIC+} derivatives and $\boldsymbol{\sigma}_S$
6. Check the predicted change in the S_{RIC+} for convergence, end macroiterations if converged
7. goto 3

The predicted change in the S_{RIC+} was found to be a more reliable convergence criteria for this algorithm than a more typical optimization criteria such as the RMS or magnitude of the S_{RIC+} gradient. When a change in the arc length of less than 10^{-2} Bohr or radians is achieved, the path has been cleaned up enough to avoid the atom collisions that can be present in the linear Cartesian path, while also satisfying the constraint conditions $\boldsymbol{\kappa}_\alpha = 0$.

This approach can also be combined with the coupling constants used in the FVRC method in order to produce an initial path that minimizes S_{RIC+}, but also travels through one or more particular geometries, such as known intermediates or guess transition state structures. In the case of bi-molecular reactions, it was previously

demonstrated[87] that interpolation of the bond order for bonds being broken or formed is an effective approach for approximating the structure at the transition state, and the methods outlined in section 6.2 can be modified to define the goal geometry only in terms of those bonds by setting $n_{act} = n_{cts}$, where n_{cts} is the number of bonds being broken or formed, and defining the goal geometry only in terms of those bonds.

6.4 Redundant Internal Coordinate VRC method

6.4.1 RIC VRE definition

As mentioned in the introduction, the VRE is the line integral of the gradient norm (eqn 6.1). Since the gradient norm has units of energy/displacement, and the tangent norm has units of displacement/dt, the integral of the gradient norm times the tangent norm over dt must have units of energy. The term in the RIC arc length formula (eqn 6.24) has units of RIC displacement/dt, so it makes sense that an RIC VRE can be constructed by incorporating a redundant internal coordinate gradient norm into S_{RIC}

$$E^q_{VRE} = \int_{t_R}^{t_P} \sqrt{\mathbf{g}_x^T \mathbf{G}_C^- \mathbf{g}_x} \sqrt{\boldsymbol{\tau}^T \left(\mathbf{G}_C + \mathbf{P}_{TR}\right) \boldsymbol{\tau}} dt \tag{6.49}$$

where $\mathbf{G}_C^-$ indicates the pseudo inverse of the inner-product G-matrix defined in section 6.3.1. $\mathbf{P}_{TR}$ only needs to be included in the tangent-norm term since it functions as a penalty on including overall translation or rotation in the path. The derivatives of $\mathbf{G}_C^-$are defined straightforwardly as

$$\frac{\partial \left[\mathbf{G}_C^-\right]_{ab}}{\partial x_i} = \sum_j \frac{\partial \left[\mathbf{B}^-\right]_{ja}}{\partial x_i} \left[\mathbf{B}^-\right]_{jb} + \left[\mathbf{B}^-\right]_{ja} \frac{\partial \left[\mathbf{B}^-\right]_{jb}}{\partial x_i} \tag{6.50}$$

where the derivative of the pseudoinverse is given by[88]

$$\frac{\partial \left[\mathbf{B}^{-}\right]_{ja}}{\partial x_i} = \sum_{k,b} - \left[\mathbf{B}^{-}\right]_{jb} \frac{\partial B_{bk}}{\partial x_i} \left[\mathbf{B}^{-}\right]_{ka} + \left[\mathbf{G}_C^{-}\right]_{jk} \frac{\partial B_{bk}}{\partial x_i} \left[\mathbf{P}_I^{\perp}\right]_{ab} + \left[\mathbf{P}_C^{\perp}\right]_{jk} \frac{\partial B_{bk}}{\partial x_i} \left[\mathbf{G}_Q^{-}\right]_{ab} \tag{6.51}$$

In order to avoid having to compute the psuedoinverse of $\mathbf{B}$ and it's derivative at every evaluation of the RIC VRE, an alternative formulation of eqn 6.49 can be used instead:

$$E_{VRE}^{q} = \int_{t_R}^{t_P} \sqrt{\mathbf{g}_q^T \mathbf{g}_q} \sqrt{\boldsymbol{\tau}^T \left(\mathbf{G}_C + \mathbf{P}_{TR}\right) \boldsymbol{\tau}} dt \tag{6.52}$$

This expression is equivalent to 6.49, but assuming that some form of surface fitting or interpolation is used to compute $\mathbf{g}_q$ directly as a function of t, the cost of computing $\mathbf{G}_C^{-}$ during the integration of 6.52 and it's derivatives may be avoided by using the chain rule, $\frac{\partial}{\partial x_i} = B_{ai} \frac{\partial}{\partial q_a}$. The derivatives of the RIC VRE may be computed in a similar fashion to the derivatives of S_{RIC+}, using eqns 6.30-6.31 and 6.37-6.39 along with the first and second x-derivatives of the RIC gradient norm

$$\frac{\partial \left|\mathbf{g}_q\right|}{\partial x_i} = \frac{1}{\left|\mathbf{g}_q\right|} \sum_a B_{ai} \left[\mathbf{H}_q \mathbf{g}_q\right]_a \tag{6.53}$$

$$\frac{\partial^2 \left|\mathbf{g}_q\right|}{\partial x_i \partial x_j} = \frac{1}{\left|\mathbf{g}_q\right|} \sum_a \left(\frac{\partial B_{ai}}{\partial x_j} \left[\mathbf{H}_q \mathbf{g}_q\right]_a + B_{ai} B_{bj} \left[\frac{\partial \mathbf{H}_q}{\partial q_b} \mathbf{g}_q\right]_a - \sum_b B_{ai} B_{bj} \left(\left[\mathbf{H}_q \mathbf{H}_q\right]_{ab} + \frac{\left[\mathbf{H}_q \mathbf{g}_q\right]_a \left[\mathbf{H}_q \mathbf{g}_q\right]_b}{\left|\mathbf{g}_q\right|^2} \right) \right) \tag{6.54}$$

6.4.2 Approximating the RIC PES

In order for the RIC VRE definition in eqn 6.52 to be useful, a method for approximating the RIC PES as a function of t is necessary. This can be accomplished efficiently by curve fitting, treating each unique element of the gradient and Hessian as a 1-dimensional function of t, and then approximating that function with a method

for interpolating a curve based upon a limited number of known points. One such interpolation method is involves using polyharmonic splines[89]. Polyharmonic splines are often used to approximate functions of more than one variable, however, they have some features that make them attractive to use here, including efficient evaluation of the fit spline, strict interpolation of data points (i.e. the fit curve is guaranteed to evaluate to the exact data at the known points), and a smooth and well behaved interpolation between points. A polyharmonic spline function $f(t)$ is evaluated as

$$f(t) = u_1 + u_2 t + \sum_{i=1} w_i \phi_r (|t - c_i|) \tag{6.55}$$

where ϕ_r is a radial basis function, the c are the values of t where the data is known (also called the centers of the interpolation), and the w and u are expansion weights determined by solving the following system of equations

$$\begin{bmatrix} \mathbf{A} & \mathbf{V}^T \\ \mathbf{V} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} \tag{6.56}$$

$$A_{ij} = \phi_r (|c_i - c_j|) \tag{6.57}$$

$$V_{1i} = 1 \tag{6.58}$$

$$V_{2i} = c_i \tag{6.59}$$

$$f_i = f(c_i) \tag{6.60}$$

where the f are the different functions being interpolated (e.g. the energy, gradient and Hessian). In the present work, the following radial function is used

$$\phi_r(r) = \begin{cases} r^3 \ln r^r & r < 1 \\ r^4 \ln r & r \geq 1 \end{cases} \tag{6.61}$$

Generally speaking, more complex target functions require more centers to achieve the same degree of accuracy. This results in a somewhat counter-intuitive result that, when the same number of centers are used, fitting the Cartesian gradient and transforming the interpolated result into internal coordinates is more accurate than fitting and interpolating the internal coordinate gradient. For this reason, only the energy, gradient and Hessian terms will be fit, rather than fitting some of the intermediate terms like $\mathbf{H}_q\mathbf{g}_q$ or $\mathbf{H}_q\mathbf{H}_q$. Additionally, when transforming the internal coordinate Hessian, additional terms that are normally neglected in geometry optimization will also be included. Normally, only the first term in eqn 6.51 is used, which is sufficient to accurately transform the Hessian in the $n_{act} \times n_{act}$ space. The final two terms are necessary to accurately transform the part of the Hessian contained in the $n_{act} \times n_{red}$ space, where $n_{red} = n_{RIC} - n_{crt} + 6$ is the number of redundant coordinates. The transformed internal coordinate Hessian expression becomes

$$\begin{aligned}[\mathbf{H}_q]_{ab} = \sum_{ij} \Bigg(& \left[\mathbf{B}^-\right]_{ia} \left[\mathbf{B}^-\right]_{jb} \left(H_{ij} + \sum_c [\mathbf{g}_q]_c \frac{\partial B_{ci}}{\partial x_j} \right) + \\ & g_i \left(\frac{\partial \left[\mathbf{B}^-\right]_{ia}}{\partial x_j} \left[\mathbf{B}^-\right]_{jb} + \frac{\partial \left[\mathbf{B}^-\right]_{ib}}{\partial x_j} \left[\mathbf{B}^-\right]_{ja} \right)\end{aligned} \tag{6.62}$$

where the plus sign in the first term is necessary to cancel out the double-counting of the $n_{act} \times n_{act}$ part in the derivative of the psuedo-inverse terms.

6.4.3 RIC-FVRC algorithm

In the algorithm outlined below, the energy, the n_{RIC} elements of the RIC gradient, and the $\frac{1}{2}n_{RIC} \times (n_{RIC}+1)$ unique elements of the RIC Hessian are fit as functions of t using polyharmonic splines at the start of each macroiteration. The PES data used in the fit is evaluated at the reactant, product, and an additional number of geometries that are equally spaced along the current path. The spline approximations are used to compute the VRE and VRE derivatives. Since the energy and gradient are fit independently of one another, the t_{ext} are optimized to maximize or minimize the energy predicted by the spline, as appropriate. Then, prior to verifying the extrema and computing the displacements for the coupling constants, the PES is evaluated at the t_{ext} to ensure that the gradient and Hessian are both accurate enough to reliably converge to the intermediates and transition states. Like with the standard FVRC method outlined in chapter **??**, the RIC-FVRC method is considered converged when the gradient at the intermediates and transition states is sufficiently small.

1. Input initial path (converged RIC-CTS path)
2. Evaluate the PES to fit the polyharmonic splines
3. Compute VRE, VRE derivatives, ϵ, and $\boldsymbol{\sigma}$
4. Locate, optimize and verify the t_{ext}, and compute the $\mathbf{q}_{ext}$ goal geometries
5. Set λ_σ and $\boldsymbol{\lambda}_\kappa$ according to 6.3.3, and set $\boldsymbol{\Delta C}$, $\mathbf{y}$, and μ_ϵ to $\mathbf{0}$
6. Begin microiterations
   (a) Compute the constraints $\kappa(\mathbf{C}+\boldsymbol{\Delta C})$ and their derivatives w.r.t. a change in the LEC
   (b) Update the $\mathbf{U}_r$ and $\boldsymbol{\Delta q}_r$ for all of the extrema
   (c) Compute $\epsilon(\mathbf{C}+\boldsymbol{\Delta C})$ and turn on optimization of μ_ϵ if $\epsilon < 0$ and $|\boldsymbol{\Delta q}_r| \approx 0$
   (d) Construct derivatives of the FVRC Lagrangian and Update $\boldsymbol{\Delta C}$, $\boldsymbol{\lambda}_\kappa$, $\mathbf{y}$, t_e and μ_ϵ

(e) Check augmented gradient and augmented displacement for convergence of microiterations, end microiterations if converged

(f) goto 5a

7. Update LEC for path, refit splines and compute the VRE, VRE derivatives and ϵ.

8. Locate, optimize and verify the t_{ext}, and compute the $\mathbf{q}_{ext}$ goal geometries

9. Check the gradient at the $\mathbf{q}_{ext}$ for convergence, and end macroiterations if converged.

10. goto 4

6.5 Results and Discussion

The methods in this chapter were implemented in Mathematica[76], using energies, gradients and Hessians computed with the Gaussian 09 electronic structure program[71]. While a more rigorous benchmarking and comparison to existing ad hoc methods is planned for a future publication, a proof of concept is provided using four example reactions (see Figure 6.1 for visualizations of the reactants and products):

MeOH	The addition of H_2 to formaldehyde to produce methanol
Ene	An ene reaction between propene and ethene to produce 1-pentane
Cope	The Cope rearrangement of 1,5-hexadiene
Bispidine	An inversion and a ring-flip starting with the axial-boat-boat configuration of N-methyl-bispidine

For each reaction, the least length RIC path was computed. In the case of the MeOH, Ene and Cope reactions, which involve the breaking and forming of bonds, the RIC-CTS path was also computed. Figure 6.2 shows the geometry for the maximum energy structure along the Cartesian interpolation, the least length RIC path, the RIC-CTS path and the final converged TS for the Ene and Cope reactions. The Cartesian path was the worst path for all reactions studied, mostly due to an unrealistic shortening of bonds not involved with the reaction, and an unrealistic lengthening of the bonds

that are breaking or forming. The least length RIC path cleans up the former but not the later, while the RIC-CTS path typically gives reasonable lengths for all bonds and differs from the actual transition state primarily by angles and torsions.

Figure 6.3 shows the compares the energy profiles for the different approximate pathways for each reaction and the final converged results from the RIC-FVRC method. In the MeOH reaction, even the Cartesian approximation is reasonably good and neither the RIC nor the RIC-CTS path offer much improvement. For the remaining reactions, however, the Cartesian interpolation results in an energy profile that's significantly worse than the other options. For the Cope and Ene reactions, the RIC-CTS path is also an improvement over the RIC. The maximum energy is slightly higher for the Cope RIC-CTS path due to the CTS approach underestimating the lengths of the bonds breaking/forming at the TS, but both the structure and the energy profile more closely resemble those of the final path than the RIC path. The RIC path for the Bispidine reaction does not involve an intermediate structure as the final path does, but it is nonetheless a clear improvement over the Cartesian path.

To demonstrate the effectiveness of the RIC-FVRC method, a single optimization was carried out for each of these reactions using 9 basis functions per coordinate, with the PES defined at the HF/3-21G level of theory. The MeOH, Ene and Cope optimizations began with the RIC-CTS path, while the Bispidine optimization began with the RIC path. In each case, a maximum step size at intermediates and transition states of 1 bohr/radian was allowed. For the MeOH and Cope reactions, a total of 9 energy/gradient/Hessian calculations was used to perform the curve fitting per iteration was sufficient to achieve convergence (RMS gradient of all transition states and intermediates of less than 10^{-4} hartree/bohr), while the Ene and Bispidine reactions required twice the density of evaluation points (19 per iteration). The total number of iterations required for convergence are:

MeOH 9 iterations
Ene 6 iterations
Cope 5 iterations
Bispidine 8 iterations

It is likely that the density of evaluations is less important than the accuracy of the energy curve, as the problem encountered with the Ene and Bispidine reactions had to do with the incorrect placement of the transition state structures, resulting in optimization steps that included too much motion along the current path and failed to systematically improve the transition state. In the present work the polyharmonic splines were fit to the energy, gradient and Hessian using uniformly sampled geometries along the path, but polyharmonic splines are also well suited to fit data with non-uniform sampling, which could improve the quality of the fit without increasing the number of PES evaluations per iteration.

The RIC-FVRC method is able to produce an approximate SDRP along with the fully converged geometries of any intermediates and transition states along the path in a small number of iterations. By approximating the PES derivatives using curve fitting techniques, this algorithmic efficiency is maintained while the per-iteration cost is reduced to something more comparable with existing ad hoc, chain of states methods. Inclusion of Hessian updating to avoid the computation of analytical Hessians should be investigated in the future, but it is expected to be non-trivial due to the role Hessian information plays in computing the VRE gradient and identifying false minima/maxima. Additionally, the development of adaptive and/or systematic approaches to fitting the PES curves could improve the accuracy of those curves while using as few evaluations of the chemical PES as possible per iteration. Once improvements such as these are included in the present methodology, the efficiency and reliability of the VRC method should compare very favorably to existing published reaction path

optimization methods.

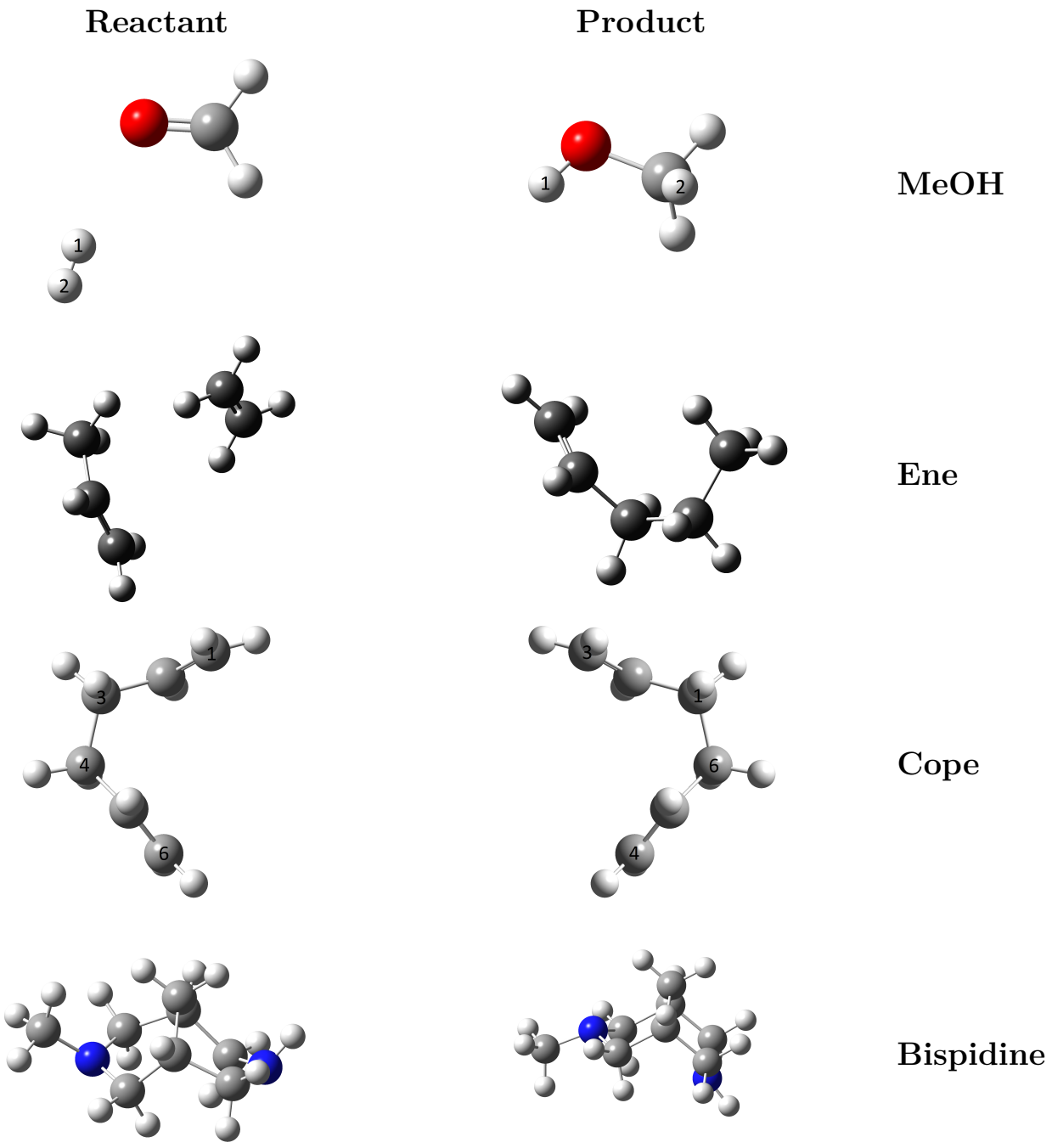


Figure 6.1: Reactant (left) and Product (right) structures. From top to bottom: MeOH, Ene, Cope, Bispidine. For the MeOH and Cope structures, key atoms have been labeled

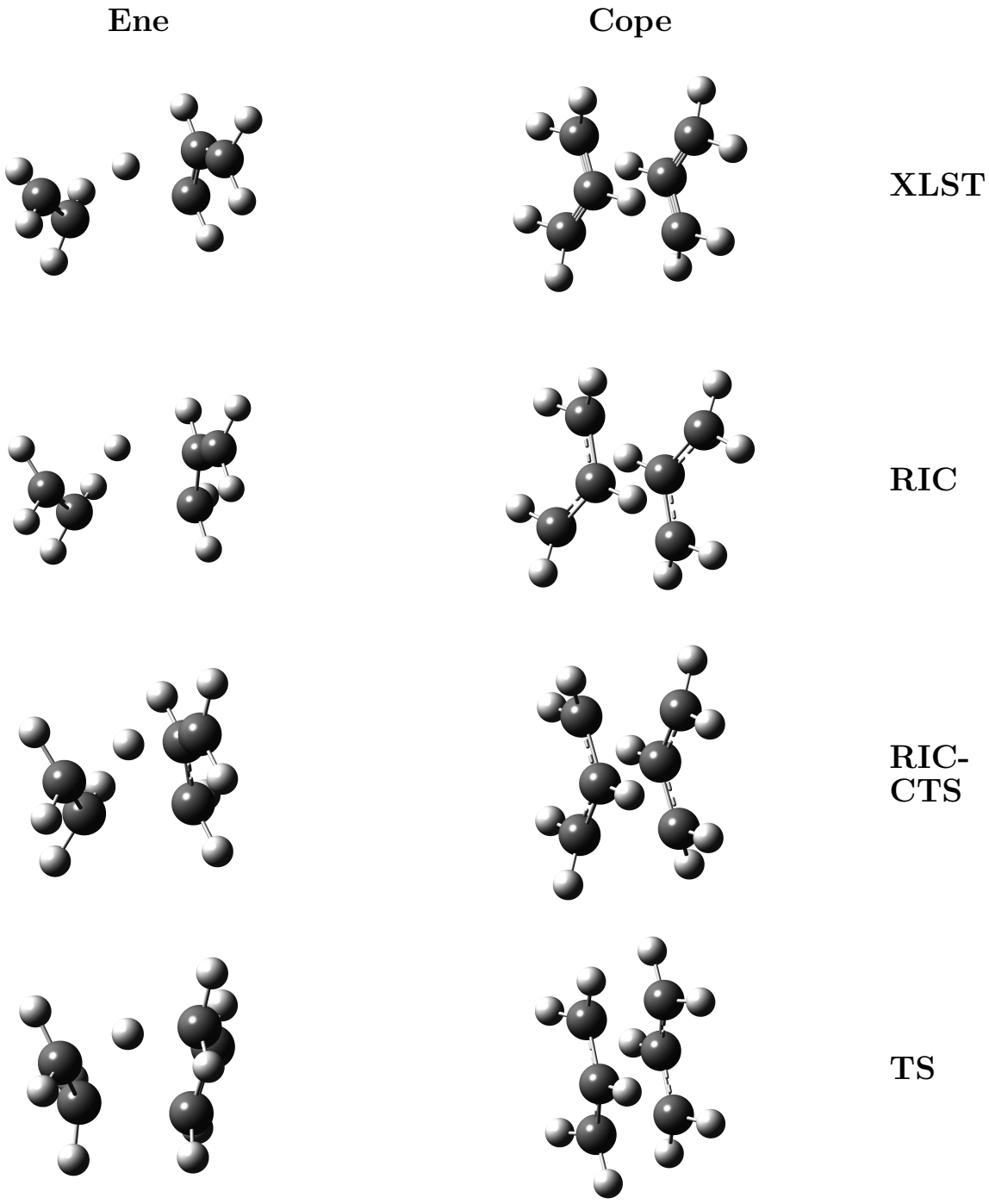


Figure 6.2: Comparison of the maximum energy structures along the path for the Ene (left) and Cope (right) reactions. From top to bottom, Cartesian interpolation (XLST), least length RIC path, RIC-CTS path, final converged path.

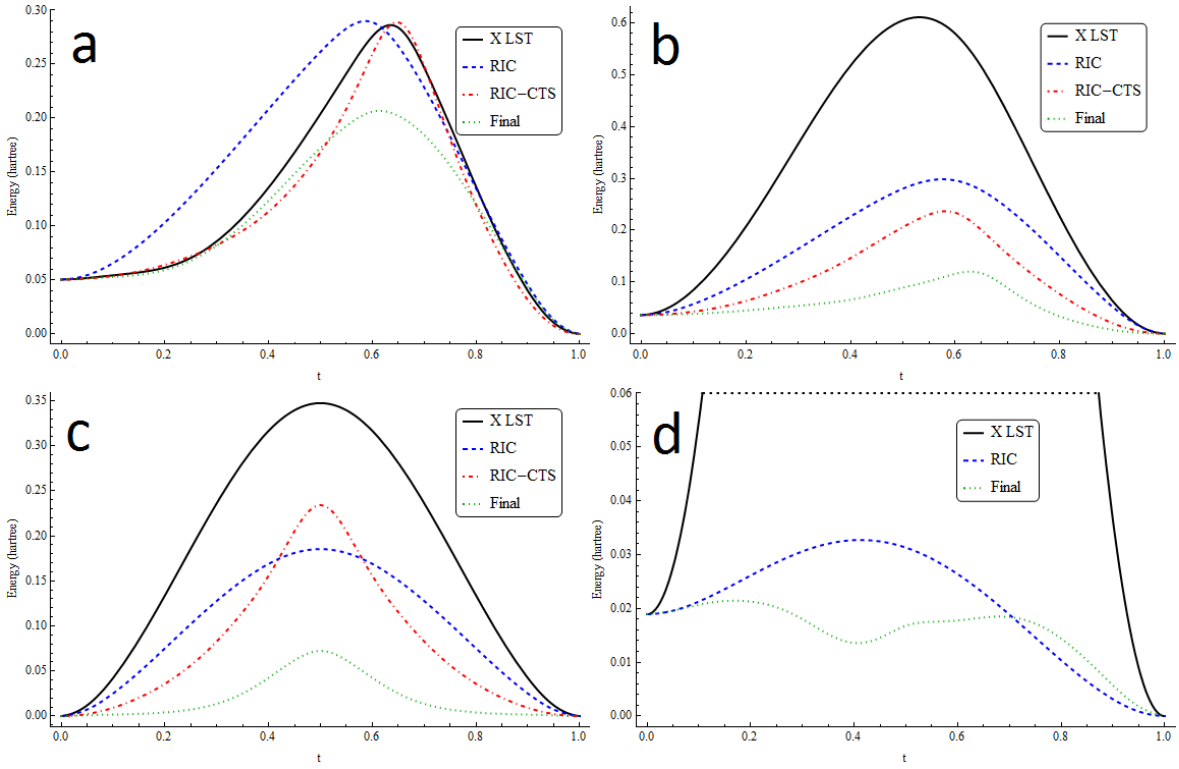


Figure 6.3: Comparison of energy profiles for the cartesian (XLST), least length RIC (RIC), RIC-CTS and final converged path. **a** MeOH, **b** Ene, **c** Cope, **d** Bispidine. The plot for Bispidine is zoomed in to show the fine structure of the RIC and final paths, the maximum along the XLST path is approximately 0.5 hartree.

CHAPTER 7

SUMMARY AND FUTURE DIRECTIONS

In this thesis, a number of new computational methods for exploring potential energy surfaces in the study of chemical reactions were introduced and examples were presented to demonstrate their efficacy. Each chapter also points to further areas for additional improvement or development.

In Chapter 2, it was demonstrated that the number of coordinates required to represent a chemical reaction could be reduced by roughly an order of magnitude without a significant impact on the accuracy in reproducing a known chemical reaction path. A reliable approach needs to be developed to perform the coordinate reduction with enough confidence that reduced space will contain a better approximation to the reaction pathway. This will likely involve including the gradient, or some form of estimated displacements, along with the information about the current path in the coordinate reduction process. This reduction could have a significant impact on the efficiency/reliability of path optimization methods.

Chapter 3 demonstrates a variety of methods for using information about the change in bonding between reactant and product structures to improve the reliability and efficiency of transition state optimization. The work presented looks exclusively at reactions consisting of a single step, but the methods discussed within this chapter may be applied to a wider variety of problems. Already, bond order interpolation has been used to improve the initial guess for reaction path optimization. Additionally, the study of reactions is often concerned with finding and comparing different pathways that may each include multiple intermediate reactions. The connectivity based methods could be expanded to create a systematic approach to studying these different reactions, by

generating multiple sequences of individual steps that would combine to produce the desired change in bonding from reactant to product, greatly simplifying the process of mapping out reaction mechanisms.

Of the minimization methods introduced in Chapter 4, the one that provided the most systematic and significant improvement was the approximately 10% reduction in optimization steps necessary for convergence achieved by the Scaled Rational Function Optimization method. The development of scaled GDIIS methods are clearly worth investigating, and the scaling approach could be applied to transition state optimization where the additional performance could have a much greater impact.

The VRC method, discussed and demonstrated in Chapters 5 and 6, perhaps has the greatest potential for additional development. The RIC-FVRC method discussed in Chapter 6 was already shown to be able to produce approximate reaction pathways, including fully optimized intermediate and transition state structures structures, in fewer than 10 iterations, with a cost of approximately 10-20 Hessian calculations per iteration. This is a significant step in the production of a new class of computational tools in the routine study of chemical reactions, but further work needs to be done for the method to enjoy widespread use. In particular, the ability to incorporate approximate and updated Hessians to reduce or eliminate the need to compute analytical Hessians will significantly decrease the computational effort required to perform an RIC-FVRC optimization. Additionally, the basis used to expand the path (quartic B-Splines), the numerical quadrature rule used (a Gauss Legendre/Curtis Clenshaw hybrid) and the method by which the potential energy surface is approximated (fitting a function of t using polyharmonic splines) are all only one possible choice out of many that could be used, and the investigation and benchmarking of alternatives could potentially not only further improve the method, but provide the foundation for a future dissertation on its own.

The computer algorithms outlined in this thesis provide a set of improved methods for exploring various features of potential energy surfaces. As with any research endeavor, the development of new methods is an ongoing process, and the results presented herein all present opportunities for further study.

REFERENCES

[1] Wales, D. Energy landscapes: applications to clusters, biomolecules and glasses (Cambridge Molecular Science) **2004**.

[2] Schlegel, H. B. Geometry optimization *WIREs Comput. Mol. Sci.* **2011** *1*, 790–809.

[3] Heidrich, D. *The reaction path in chemistry: current approaches and perspectives* volume 16 Springer Science & Business Media **2013**.

[4] Elber, R.; Karplus, M. A method for determining reaction paths in large molecules: application to myoglobin *Chem. Phys. Lett.* **1987** *139*, 375–380.

[5] Burger, S. K.; Yang, W. Quadratic string method for determining the minimum-energy path based on multiobjective optimization *J. Chem. Phys.* **2006** *124*, 054109.

[6] Sheppard, D.; Terrell, R.; Henkelman, G. Optimization methods for finding minimum energy paths *J. Chem. Phys.* **2008** *128*, 134106.

[7] Schatz, G. C. The analytical representation of electronic potential-energy surfaces *Rev. Mod. Phys.* **1989** *61*, 669.

[8] Ischtwan, J.; Collins, M. A. Molecular potential energy surfaces by interpolation *J. Chem. Phys.* **1994** *100*, 8080–8088.

[9] Doubleday, C.; Bolton, K.; Peslherbe, G. H.; Hase, W. L. Direct dynamics simulation of the lifetime of trimethylene *J. Am. Chem. Soc.* **1996** *118*, 9922–9931.

[10] Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. The calculation of ab initio molecular geometries: efficient optimization by natural internal coordinates and empirical correction by offset forces *J. Am. Chem. Soc.* **1992** *114*, 8191–8201.

[11] Pulay, P.; Fogarasi, G. Geometry optimization in redundant internal coordinates *J. Chem. Phys.* **1992** *96*, 2856–2860.

[12] Baker, J.; Chan, F. The location of transition states: A comparison of Cartesian, Z-matrix, and natural internal coordinates *J. Comput. Chem.* **1996** *17*, 888–904.

[13] Kato, S.; Morokuma, K. Potential energy characteristics and energy partitioning in chemical reactions: Abinitio MO study of four-centered elimination reaction $CH_3CH_2F \rightarrow CH_2{=}CH_2 + HF$ *J. Chem. Phys.* **1980** *73*, 3900–3914.

[14] Konkoli, Z.; Kraka, E.; Cremer, D. Unified Reaction Valley Approach Mechanism of the Reaction $CH_3 + H_2 \rightarrow CH_4 + H$ *J. Phys. Chem. A* **1997** *101*, 1742–1757.

[15] Jolliffe, I. *Principal component analysis* Wiley Online Library **2002**.

[16] Jensen, F.; Palmer, D. S. Harmonic vibrational analysis in delocalized internal coordinates *J. Chem. Theory Comput.* **2010** *7*, 223–230.

[17] Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. Principal component analysis and long time protein dynamics *J. Phys. Chem.* **1996** *100*, 2567–2572.

[18] Palmer, D. S.; Jensen, F. Predicting large-scale conformational changes in proteins using energy-weighted normal modes *Proteins: Struct., Funct., Bioinf.* **2011** *79*, 2778–2793.

[19] Fukui, K. The path of chemical reactions-the IRC approach *Acc. Chem. Res.* **1981** *14*, 363–368.

[20] Gonzalez, C.; Schlegel, H. B. An improved algorithm for reaction path following *J. Chem. Phys.* **1989** *90*, 2154–2161.

[21] Gonzalez, C.; Schlegel, H. B. Reaction path following in mass-weighted internal coordinates *J. Phys. Chem.* **1990** *94*, 5523–5527.

[22] Broyden, C. G. The convergence of single-rank quasi-Newton methods *Math. Comput.* **1970** *24*, 365.

[23] Fletcher, R. A new approach to variable metric algorithms *Comput. J.* **1970** *13*, 317–322.

[24] Goldfarb, D. A family of variable-metric methods derived by variational means

Math. Comput. **1970** *24*, 23.

[25] Shanno, D. F. Conditioning of quasi-Newton methods for function minimization *Math. Comput.* **1970** *24*, 647.

[26] Powell, M. J. *A new algorithm for unconstrained optimization* UKAEA **1970**.

[27] Murtagh, B. A. Computational experience with quadratically convergent minimisation methods *Comput. J.* **1970** *13*, 185–194.

[28] Bofill, J. M. Updated Hessian matrix and the restricted step method for locating transition structures *J. Comput. Chem.* **1994** *15*, 1–11.

[29] Schlegel, H. B. Estimating the hessian for gradient-type geometry optimizations *Theor. Chim. Acta.* **1984** *66*, 333–340.

[30] Császár, P.; Pulay, P. Geometry optimization by direct inversion in the iterative subspace *Journal of Molecular Structure* **1984** *114*, 31–34.

[31] Hratchian, H. P.; Schlegel, H. B. Accurate reaction paths using a Hessian based predictor–corrector integrator *J. Chem. Phys.* **2004** *120*, 9918–9924.

[32] Halgren, T. A.; Lipscomb, W. N. The synchronous-transit method for determining reaction pathways and locating molecular transition states *Chem. Phys. Lett.* **1977** *49*, 225–232.

[33] Crehuet, R.; Bofill, J. M. The reaction path intrinsic reaction coordinate method and the Hamilton–Jacobi theory *J. Chem. Phys.* **2005** *122*, 234105.

[34] Quapp, W. Chemical reaction paths and calculus of variations *Theor. Chem. Acc.* **2008** *121*, 227–237.

[35] Hratchian, H.; Schlegel, H. Using Hessian updating to increase the efficiency of a Hessian based predictor-corrector reaction path following method *J. Chem. Theory Comput.* **2005** *1*, 61–69.

[36] Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; et al. Gaussian 09, Rev.

B. 01 *Gaussian Inc., Wallingford CT* **2010**.

[37] Mathematica Version 7.0 Wolfram Research, Inc. Champaign, IL **2008**.

[38] Heathcock, C. H.; Mahaim, C.; Schlecht, M. F.; Utawanit, T. A synthetic approach to the quassinoids *J. Org. Chem.* **1984** *49*, 3264–3274.

[39] Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks III, C. L.; Pople, J. A. Theoretical study of blocked glycine and alanine peptide analogs *J. Am. Chem. Soc.* **1991** *113*, 5989–5997.

[40] Stewart, J. J. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements *J. Mol. Model.* **2007** *13*, 1173–1213.

[41] Sonnenberg, J. L.; Wong, K. F.; Voth, G. A.; Schlegel, H. B. Distributed Gaussian valence bond surface derived from ab initio calculations *J. Chem. Theory Comput.* **2009** *5*, 949–961.

[42] Niu, S.; Hall, M. B. Theoretical studies on reactions of transition-metal complexes *Chem. Rev.* **2000** *100*, 353–406.

[43] Zhou, J.; Tao, P.; Fisher, J. F.; Shi, Q.; Mobashery, S.; Schlegel, H. B. QM/MM studies of the matrix metalloproteinase 2 (MMP2) inhibition mechanism of (S)-SB-3CT and its oxirane analogue *J. Chem. Theory Comput.* **2010** *6*, 3580–3587.

[44] Perdew, J. P.; Ziesche, P.; Eschrig, H. *Electronic structure of solids '91* volume 11 Akademie Verlag, Berlin **1991**.

[45] Perdew, J. P. Generalized gradient approximations for exchange and correlation: A look backward and forward *Physica B: Condensed Matter* **1991** *172*, 1–6.

[46] Perdew, J. P.; Chevary, J.; Vosko, S.; Jackson, K. A.; Pederson, M. R.; Singh, D.; Fiolhais, C. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation *Phys. Rev. B* **1992** *46*, 6671.

[47] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange

J. Chem. Phys. **1993** *98*, 5648–5652.

[48] Wales, D. *Energy landscapes: applications to clusters, biomolecules and glasses* Cambridge University Press **2003**.

[49] Olsen, R.; Kroes, G.; Henkelman, G.; Arnaldsson, A.; Jónsson, H. Comparison of methods for finding saddle points without knowledge of the final states *J. Chem. Phys.* **2004** *121*, 9776–9792.

[50] Sharada, S. M.; Bell, A. T.; Head-Gordon, M. A finite difference Davidson procedure to sidestep full ab initio hessian calculation: Application to characterization of stationary points and transition state searches *J. Chem. Phys.* **2014** *140*, 164115.

[51] Zeng, Y.; Xiao, P.; Henkelman, G. Unification of algorithms for minimum mode optimization *J. Chem. Phys.* **2014** *140*, 044115.

[52] Fischer, T. H.; Almlof, J. General methods for geometry and wave function optimization *J. Phys. Chem.* **1992** *96*, 9768–9774.

[53] Wittbrodt, J. M.; Schlegel, H. B. Estimating stretching force constants for geometry optimization *TheoChem* **1997** *398*, 55–61.

[54] Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. Search for stationary points on surfaces *J. Phys. Chem.* **1985** *89*, 52–57.

[55] Bofill, J. M.; Anglada, J. M. Finding transition states using reduced potential-energy surfaces *Theor. Chem. Acc.* **2001** *105*, 463–472.

[56] Burger, S. K.; Ayers, P. W. Dual grid methods for finding the reaction path on reduced potential energy surfaces *J. Chem. Theory Comput.* **2010** *6*, 1490–1497.

[57] Burger, S. K.; Ayers, P. W. Methods for finding transition states on reduced potential energy surfaces *J. Chem. Phys.* **2010** *132*, 234110.

[58] Peng, C.; Bernhard Schlegel, H. Combining Synchronous Transit and Quasi-Newton Methods to Find Transition States *Israel J. Chem.* **1993** *33*, 449–454.

[59] Henkelman, G.; Jonsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives *J. Chem. Phys.* **1999** *111*, 7010.

[60] Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Efficient exploration of reaction paths via a freezing string method *J. Chem. Phys.* **2011** *135*, 224108.

[61] Plessow, P. Reaction Path Optimization without NEB Springs or Interpolation Algorithms *J. Chem. Theory Comput.* **2013** *9*, 1305–1310.

[62] Mallikarjun Sharada, S.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Automated transition state searches without evaluating the Hessian *J. Chem. Theory Comput.* **2012** *8*, 5166–5174.

[63] Zimmerman, P. Reliable Transition State Searches Integrated with the Growing String Method *J. Chem. Theory Comput.* **2013** *9*, 3043–3050.

[64] Birkholz, A. B.; Schlegel, H. B. Coordinate reduction for exploring chemical reaction paths *Theor. Chem. Acc.* **2012** *131*, 1–8.

[65] Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states *J. Comput. Chem.* **1996** *17*, 49–56.

[66] Ess, D. H. Transition-Structure Catalog of Organic Reactions *J. Chem. Educ.* **2012** *89*, 817–818.

[67] Pauling, L. Atomic radii and interatomic distances in metals *J. Am. Chem. Soc.* **1947** *69*, 542–553.

[68] Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard III, W.; Skiff, W. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations *J. Am. Chem. Soc.* **1992** *114*, 10024–10035.

[69] Bell, S.; Crighton, J. S. Locating transition states *J. Chem. Phys.* **1984** *80*, 2464–2475.

[70] May, J. W.; Lehner, J. D.; Frisch, M. J.; Li, X. Transition State Search Using a Guided Direct Inversion in the Iterative Subspace Method *J. Chem. Theory Comput.* **2012** *8*, 5175–5179.

[71] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 gaussian Inc. Wallingford CT 2009.

[72] Vosko, S.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis *Can. J. Phys.* **1980** *58*, 1200–1211.

[73] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density *Phys. Rev. B* **1988** *37*, 785.

[74] Stephens, P.; Devlin, F.; Chabalowski, C.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields *J. Phys. Chem.* **1994** *98*, 11623–11627.

[75] Löwdin, P.-O. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals *J. Chem. Phys.* **1950** *18*, 365–375.

[76] Mathematica Version 9.0 Wolfram Research, Inc. Champaign, IL **2013**.

[77] Ayala, P. Y.; Schlegel, H. B. A combined method for determining reaction paths, minima, and transition state geometries *J. Chem. Phys.* **1997** *107*, 375–384.

[78] Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points *J. Chem. Phys.* **2000** *113*, 9978–9985.

[79] Ren, W.; Vanden-Eijnden, E.; et al. String method for the study of rare events *Phys. Rev. B* **2002**.

[80] Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods *J. Chem. Phys.* **2004** *120*, 7877–7886.

[81] De Boor, C. A practical guide to splines *Mathematics of Computation* **1978**.

[82] Dash, R. B.; Das, D. in A mixed quadrature rule by blending Clenshaw-Curtis and Gauss-Legendre quadrature rules for approximation of real definite integrals in adaptive environmentvolume 1.

[83] Peng, C.; Schlegel, H. B. Combining synchronous transit and quasi-Newton methods to find transition states *Israel Journal of Chemistry* **1993** *33*, 449–454.

[84] Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure *Theoretica chimica acta* **1979** *53*, 75–93.

[85] Lennard-Jones, J. E. in On the determination of molecular fields.volume 106 The Royal Society 441–462.

[86] Wales, D.; Doye, J.; Dullweber, A.; Hodges, M.; Naumkin, F.; Calvo, F.;

Hernández-Rojas, J.; Middleton, T. The Cambridge cluster database *URL http://doye.chem.ox.ac.uk/networks/LJn.html* **2002**.

[87] Birkholz, A. B.; Schlegel, H. B. Using bonding to guide transition state optimization *J. Comput. Chem.* **2015** *36*, 1157–1166.

[88] Golub, G. H.; Pereyra, V. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate *SIAM J. Numer. Anal.* **1973** *10*, 413–432.

[89] Madych, W. R.; Nelson, S. Polyharmonic cardinal splines *J. Approx. Theory* **1990** *60*, 141–156.

ABSTRACT

DEVELOPMENT OF NEW ALGORITHMS FOR EXPLORING THE POTENTIAL ENERGY LANDSCAPE OF CHEMICAL REACTIONS

by

ADAM BENJAMIN BIRKHOLZ

December 2015

Advisor: Professor H. Bernhard Schlegel

Major: Chemistry (Physical)

Degree: Doctor of Philosophy

The research presented in this dissertation is divided into 5 chapters. In Chapter 2, a method for reducing the number of coordinates required to accurately reproduce a known chemical reaction pathway by applying principal component analysis to a number of geometries along the pathway (expressed in either Cartesian coordinates or redundant internal coordinates) is described and applied to 9 example reactions. Chapter 3 introduces new methods for estimating the structure of and optimizing transition states by utilizing information about the atomic bonding in the reactants and products. These methods are then benchmarked against a standard transition state optimization approach utilizing a test set of 20 reactions, with energies computed at both semiempirical and density functional theory levels of theory. Chapter 4 is a collection of 3 new, alternative approaches (Flowchart

Hessian updating, Scaled Rational Function Optimization, Quasi-Rotation coordinate propagation), to handling aspects of a typical Quasi-Newton minimization. These new approaches are then compared to their standard counterparts by optimizing a set of 20 molecules using either ab initio or density functional theory potential energy surfaces.

The final two chapters of this thesis focus on the development of a new path optimization framework, the Variational Reaction Coordinate (VRC) method. This framework seeks to improve upon the "chain of states" methods, which minimize the energy of a series of structures while using constraints, fictitious forces or reparameterization schemes to maintain the distribution of points along the path. In the VRC method, a functional representing the energy of the entire reaction is minimized by varying the expansion coefficients of a continuous function used to represent the reaction path. In Chapter 5, an algorithm is outlined along with the discussion and application of constraints and coupling terms that may be used to improve the efficiency and reliability of the method, with analytical test surfaces used to demonstrate the method's performance. Chapter 6 focuses on the inclusion of redundant internal coordinates and methods for approximating the potential energy surface into the VRC framework, in order to reduce the per-iteration computational cost of the VRC method to something comparable to the "chain of states" approaches so that it may be practically applied to the study of reactions using high accuracy density functional theory and ab initio potential energy surfaces.

AUTOBIOGRAPHICAL STATEMENT

Adam Benjamin Birkholz

Born on November 24th, 1984 in Minneapolis, Minnesota, USA

EDUCATION

- Doctorate of Philosophy in Physical Chemistry, September 2015
  Wayne State University, Detroit, Michigan
  Advisor: Professor H. Bernhard Schlegel
- Baccalaureate of Science in Chemistry, December 2005
  University of Minnesota, Twin Cities campus, Minneapolis, Minnesota

WORK EXPERIENCE

- Quality Control Chemist, February 2007 – August 2008
  NeoMPS, Inc., San Diego, CA
- Lab Assistant, January 2006 – August 2006
  Metropolitan Council, St. Paul, MN

PRESENTATIONS

- Making Ends Meet: Reaction Path Optimization on a Budget
  Spring 2011 ACS National Meeting, Anaheim, CA
  Midwest Theory Conference 2011 (Notre Dame), South Bend, IN
- Development of a Variational Reaction Coordinate Method
  Fall 2012 ACS National Meeting, Philadelphia, PA

PEER-REVIEWED PUBLICATIONS

- Coordinate Reduction for Exploring Chemical Reaction Paths, **A. Birkholz,** H.B. Schlegel, Theor. Chem. Acc. 2012 *131*, 1-8 (doi: 10.1007/s00214-012-1170-6)
- Using Bonding to Guide Transition State Optimization, **A. Birkholz**, H.B. Schlegel, J. Comput. Chem. 2015 *36(15)*, 1157-1166 (doi: 10.1002/jcc.23910)